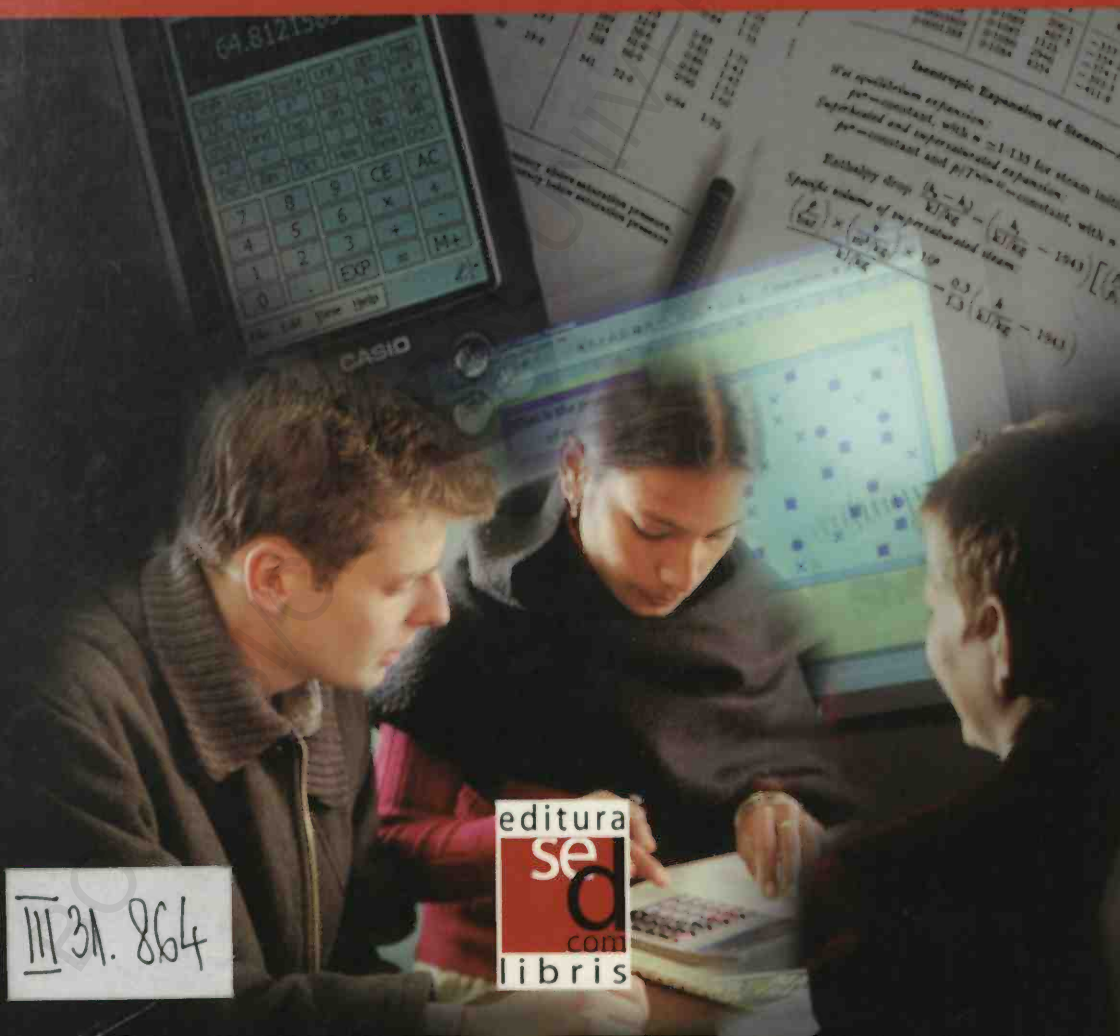
Elisabeta JABA

Dănuţ JEMNA

# Econometrie




editura sedcom libris

III 31. 864

BIBL. CENTR. UNIV.
„M. EMINESCU" IAŞI
III 31. 864

BCU IAŞI/CENTRAL UNIVERSITY LIBRARY

# Econometrie

BCU "ASICENTRAL UNIVERSITY LIBRARY

**Redactor: Petru RADU**
Revizia textelor: Alina HUCAI
Concepția și realizarea tehnică a copertei: Mona Eliza VELEȘCU
Tehnoredactare și revizia tehnică a copertei: Alina ASANDEI
Pre-press: Ciprian TOȘCARIU

*Tehnoredactarea computerizată aparține autorilor.*

**Descrierea CIP a Bibliotecii Naționale a României**

**JABA, ELISABETA**
**Econometrie** / Elisabeta Jaba, Dănuț Jemna.
– Iași: Sedcom Libris, 2006
Bibliogr.
**I.S.B.N. (10): 973-670-208-1**
**I.S.B.N. (13): 978-973-670-208-2**

I. Jemna, Dănuț

330.43

**Editura Sedcom Libris este acreditată de Consiliul Național al Cercetării Științifice din Învățământul Superior (C.N.C.S.I.S.).**

**Copyright © 2006, SEDCOM LIBRIS**
Toate drepturile asupra prezentei ediții sunt rezervate Editurii Sedcom Libris, Iași. Reproducerea parțială sau integrală a textelor, prin orice mijloc, precum și a graficii copertei, fără acordul scris al Editurii Sedcom Libris, este interzisă și se va pedepsi conform legislației în vigoare.

*Adresa Editurii*: Șos. Moara de Foc nr. 4, cod 700527, Iași, România
*Contact Editura*:
Tel.: +40.232.242.877; 234.582; 0742.76.97.72; fax: +40.232.233.080
www.sedcom.ro; e-mail: editurasedcomlibris@yahoo.com

BIBL. CENTR. UNIV. M. EMINESCU IAȘI
BIBLIOTECA CENTRALĂ UNIVERSITY LIBRARY

Elisabeta JABA

Dănuţ JEMNA

443 509

# Econometrie

444795
B.C.U. IASI

ALE

editura sedcom libris

BCU IASI CENTRAL UNIVERSITY LIBRARY

**Referenți științifici:**

**Prof. univ. dr. Elena-Maria BIJI**

**Prof. univ. dr. Constantin MITRUȚ**

BCU IASI CENTRAL UNIVERSITY LIBRARY

9. APR. 2009

# CUPRINS

CUVÂNT ÎNAINTE ..... 7

CAPITOLUL I
NOŢIUNI FUNDAMENTALE ..... 9
1.1. CE ESTE ECONOMETRIA? ..... 9
1.2. REPERE ISTORICE ..... 11
1.3. CONCEPTE ..... 12
1.4. DEMERS METODOLOGIC ..... 16
1.5. DATE STATISTICE ÎN CERCETAREA ECONOMETRICĂ .. 17
1.6. NOTAŢII ..... 18

CAPITOLUL II
MODELE DE REGRESIE SIMPLĂ ..... 19
2.1. MODELUL LINIAR SIMPLU ..... 19
2.1.1 Formularea problemei. Exemple din economie ..... 20
2.1.2 Model şi ipoteze ..... 20
2.1.3 Estimarea parametrilor modelului ..... 24
2.1.4 Testarea parametrilor modelului ..... 29
2.1.5 Coeficienţi de corelaţie ..... 33
2.1.6 Regresia liniară simplă în SPSS ..... 39
2.2. MODELUL NELINIAR SIMPLU ..... 45
2.2.1 Formularea problemei şi exemple din economie ..... 45
2.2.2 Modele liniarizabile ..... 46
2.2.3 Modele polinomiale ..... 48

CAPITOLUL III
MODELE DE REGRESIE MULTIPLĂ ..... 55
3.1. MODELUL LINIAR MULTIPLU ..... 55
3.1.1 Prezentare model şi ipoteze ..... 56
3.1.2 Estimarea parametrilor modelului ..... 56
3.1.3 Testarea parametrilor modelului ..... 57
3.1.4 Coeficienţi de corelaţie ..... 60
3.1.5 Regresia multiplă în SPSS ..... 67
3.2. MODELE NELINIARE MULTIPLE ..... 76

3.2.1 Prezentarea problemei. Exemple din economie ..... 77
3.2.2 Modele liniarizabile ..... 77
3.2.3 Modele polinomiale ..... 82

CAPITOLUL IV
TESTAREA IPOTEZELOR MODELULUI CLASIC DE REGRESIE ... 83
4.1. NORMALITATEA ERORILOR ..... 84
4.2. HOMOSCEDASTICITATEA ..... 85
4.3. LIPSA AUTOCORELĂRII ERORILOR ..... 89
4.4. LIPSA COLINIARITĂŢII VARIABILELOR INDEPENDENTE ..... 96
4.5. TESTAREA IPOTEZELOR ÎN SPSS ..... 99

CAPITOLUL V
MODELE SPECIALE ..... 111
5.1. VARIABILE DUMMY ..... 112
5.2. MODELE ANOVA ..... 112
5.3. MODELE ANCOVA ..... 117
5.4.APLICAŢII ALE MODELELOR ANCOVA ..... 124

CAPITOLUL VI
MODELE PENTRU SERII DE TIMP ..... 127
6.1. COMPONENTELE UNEI SERII DE TIMP ..... 128
6.2. ESTIMAREA TRENDULUI ..... 132
6.3. AJUSTAREA SERIILOR SEZONIERE ..... 137
6.4.METODE MODERNE DE ANALIZĂ A SERIILOR DE TIMP ..... 144
6.4.1 Procese stochastice ..... 145
6.4.2 Tipuri de procese stochastice ..... 149
6.4.3 Prognoză prin metoda Box-Jenkins ..... 152
6.5. EXEMPLE ÎN SPSS ..... 153

ANEXĂ ..... 171
BIBLIOGRAFIE ..... 193
TABELE PROBABILISTE ..... 203

# CUVÂNT ÎNAINTE

*Clasicii economiei politice au făcut deseori previziuni asupra viitorului economiei mondiale sau al unei țări anume. Sub aspectul dimensiunii și problematicii, dinamica vieții economice actuale a depășit așteptările și previziunile economiștilor din trecut. În plan economic, asistăm la un proces de globalizare. Economia a devenit un sistem deschis, dinamic și predispus la restructurări importante.*

*Pentru cercetătorii din domeniul economic, acest proces de dezvoltare reprezintă în permanență o sursă de provocări, un ansamblu de probleme de rezolvat și un cadru empiric de evaluare și dezvoltare a teoriilor economice, a metodologiilor de cercetare și a paradigmelor de gândire.*

*Econometria este produsul acestei relații dinamice dintre realitatea economică și munca de cercetare. Scop principal al econometriei este acela de a explica modul în care se intercondiționează și se produc fenomenele economice.*

*Premisele dezvoltării econometriei ca disciplină științifică sunt legate de dezvoltarea teoriei inferenței statistice, a teoriei macroeconomice, a instrumentelor de evaluare a rezultatelor economice sub forma unor agregate, precum și de creșterea cererii de cercetări științifice asupra realității economice. La acestea am putea adăuga perfecționarea sistemelor de calcul informatic și reducerea considerabilă a timpului alocat prelucrării datelor statistice.*

*Ca răspuns la dinamica realității economice și la nevoia de cunoaștere a acesteia, econometria se regăsește în planurile de învățământ ale facultăților de economie din aproape orice universitate din lume. Începând cu anul universitar 2005-2006, prin adoptarea sistemului de învățământ european propus la Bologna, în cea mai mare parte a facultăților de economie din România, econometria apare ca o disciplină fundamentală în formarea economiștilor cu studii superioare.*

*Lucrarea de față este destinată studenților și specialiștilor din domeniul economic. Se constituie ca un corp de elemente de bază necesare însușirii și aplicării procesului de modelare a fenomenelor economice.*

BCU "MIHAI EMINESCU" IAȘI UNIVERSITY LIBRARY

*Structura lucrării confirmă o abordare orientată spre înțelegerea demersului modelării econometrice pe tipuri de modele.*

*În primul capitol, alături de definiție și un scurt istoric, sunt prezentate noțiunile fundamentale necesare în cercetarea econometrică: conceptul de model și componentele acestuia, demersul metodologic, tipurile de date folosite în cercetarea econometrică etc.*

*În capitolul II, sunt prezentate modelele econometrice simple, numite și modele de regresie simplă, grupate în modele liniare și neliniare. Pentru fiecare tip de model se tratează problemele specifice demersului metodologic: formularea problemei, identificarea modelului, estimarea și testarea parametrilor modelului.*

*Capitolul III este rezervat modelelor de regresie multiplă, respectând structura și problematica elaborată în capitolul anterior.*

*În capitolul IV, sunt supuse analizei principalele restricții sau ipoteze pe care trebuie să le respecte un model econometric pentru a îndeplini condițiile de calitate cerute de metoda statistică.*

*Capitolul V este dedicat unei categorii speciale de modele, modele cu variabile independente dummy.*

*Ultimul capitol este destinat modelelor specifice seriilor de timp. Sunt prezentate problemele analizei unei serii de timp pe componente și ale modelării unei serii de timp, în scopul realizării de prognoze.*

*Lucrarea se încheie cu o anexă care conține elemente de teoria probabilităților, o bibliografie selectivă din domeniu de specialitate și un set de tabele probabiliste utile în etapele de estimare și testare a modelelor.*

*Pentru îmbunătățirea demersului întreprins, întrebările și sugestiile dumneavoastră vor fi de un real folos.*

*Un gând de mulțumire pentru cei toți care au studiat și au transpus în lucrări ideile lor asupra domeniului, dându-ne posibilitatea să înțelegem un tărâm fascinant, într-o continuă dinamică a cunoașterii. De asemenea, calde mulțumiri colegilor, alături de care, prin dezbateri, am descoperit noi dimensiuni.*

*Noiembrie 2006* *Autorii*

# Capitolul 1

# NOŢIUNI FUNDAMENTALE

1. Ce este econometria?
2. Repere istorice
3. Concepte
4. Demers metodologic
5. Date statistice în cercetarea econometrică
6. Notaţii

B.C.U. "M. EMINESCU" IAŞI

În acest capitol, sunt prezentate noţiunile fundamentale ale econometriei ca disciplină ştiinţifică. Sunt vizate: definiţia econometriei, obiectul, metoda şi scopul acesteia, câteva repere istorice ale dezvoltării econometriei, conceptele de bază, demersul metodologic şi notaţiile utilizate.

## 1.1. Ce este econometria?

Termenul *econometrie*[1] a fost introdus de economistul şi statisticianul norvegian R. Frisch, în anul 1926, prin analogie cu termenul „biometrie" utilizat de Galton şi Pearson. Aşa cum prin „biometrie" se

[1] În limba engleză, *econometrics*, creat pe baza termenilor *economics metrics*, *economics measurement*.

desemnau cercetările biologice cu ajutorul statisticii şi matematicii, econometria avea să însemne studiul economiei cu ajutorul acestor ştiinţe fundamentale.

### Definiţie

În sens etimologic, *econometria* înseamnă măsurarea sau exprimarea cantitativă a fenomenelor economice. Lansarea şi semnificaţiile termenului sunt legate de Societatea de Econometrie[2].

Econometria este o disciplină care s-a conturat ca o sinteză între economie, matematică şi statistică.

Plecând de la teoriile economice, pe baza datelor reale din economie, econometria construieşte modele (expresii cantitative) pentru exprimarea legăturilor dintre fenomenele economice studiate şi variaţia lor în timp. Pe baza modelelor, folosind procedeele de inferenţă statistică, econometria realizează predicţii asupra fenomenelor analizate.

Ca disciplină, econometria se împarte, în funcţie de obiectul cercetării, în două componente: analiza legăturilor dintre fenomenele economice şi analiza seriilor de timp.

### Obiect

Aria de studiu a econometriei este realitatea economică privită ca un ansamblu de relaţii şi intercondiţionări. Econometria studiază legăturile dintre fenomenele economice, dintre diferite componente ale economiei în ansamblul său, precum şi variaţia în timp a fenomenelor.

### Metodă

Econometria contribuie la cunoaşterea realităţii economice prin modul său specific de a surprinde cantitativ legăturile dintre fenomenele economice cu ajutorul unui instrument specific: modelul econometric. Procedeele specifice econometriei sunt estimarea şi testarea modelelor econometrice, prin care se surprind relaţiile dintre fenomenele economice reale.

---

[2] Fondatorii Societăţii de Econometrie (1930) au definit econometria astfel: „the quantitative analysis of actual economic phenomena based on the concurrent development of theory and observation, related by appropriate methods of inference – Samuelson, P., Koopmans, T. and Stone, R., Report of the Evaluative Committee for Econometrica, *Econometrica*, 1954, p. 142.

### Scop

Econometria are un scop explorativ şi explicativ, vizând descoperirea de noi legi de manifestare a fenomenelor din economie.

## 1.2. Repere istorice

Econometria este o disciplină ştiinţifică relativ nouă, dezvoltată începând cu mijlocul secolului trecut. Totuşi, primele încercări de a cuantifica şi exprima cantitativ relaţiile dintre fenomenele reale sunt mult mai vechi şi datează din secolul al XVII-lea.

### Şcoala Aritmeticii politice engleze

Geneza econometriei ar putea fi considerată la începuturile secolului al XVII-lea când englezul W. Petty pune bazele „aritmeticii politice". În cadrul şcolii create, se foloseau sistematic fapte şi cifre pentru elaborarea unor studii legate de populaţie, finanţe, comerţ exterior sau impozitare.

### Laboratoarele biometrice engleze

La sfârşitul secolului al XIX-lea şi începutul secolului al XX-lea, în Anglia se desfăşura o activitate ştiinţifică remarcabilă de cercetare a legilor naturii şi a geneticii umane. Printre figurile ilustre ale acestei şcoli se numără F. Galton, K. Pearson, R.A. Fisher, F.Y. Edgeworth, ale căror lucrări fundamentale au contribuit la dezvoltarea metodelor de analiză a legăturilor dintre variabile.

### Societatea de econometrie

La 29 decembrie 1930, la Cleveland (S.U.A.) a fost întemeiată „Societatea de Econometrie", instituţie care a creat şi promovat termenul de „econometrie".

Dintre membrii societăţii, menţionăm cele mai importante figuri: Irving Fisher, R. A. Fisher (matematician şi biolog, care a dezvoltat analiza dispersională), Jan Timbergen (fizician olandez), Trygve Haavelmo, R. Frisch (primul preşedinte al societăţii) ş.a.

Societatea de Econometrie lansează, în anul 1933, prima revistă de econometrie, *Econometrica*, avându-l ca redactor-şef pe R. Frisch. Revista

va juca un rol important în dezvoltarea econometriei prin popularizarea cercetărilor din domeniu şi prin promovarea noilor specialişti.

### Mari gânditori ai secolului al XX-lea

Econometria se dezvoltă prin contribuţia unor cercetători importanţi, din diferite direcţii ale cercetării:

- producţie: Cobb C.W. şi Douglas P.H.;
- cererea de consum: K. Schultz, P.A. Samuelson;
- teoriile economice şi construirea modelelor: J. Timbergen, T. Haavelmo, R. Frisch, L. R. Klein, H. Theil;
- studiul riscului şi incertitudinii în economie, modele macroeconomice: J.M. Keynes.

## 1.3 Concepte

În cercetarea econometrică, se utilizează o serie de concepte, noţiuni şi termeni specifici: *model, variabile, parametri, estimator, estimaţii* etc.

### Modelul econometric

Conceptul de *model econometric* s-a dezvoltat pe baza teoriei economice, în care este definit conceptul de *model economic*, şi a matematicii economice, care defineşte şi operează cu modelul matematic.

Modelul este o schemă simplificată a realităţii. Are rolul de a explica realitatea studiată în dimensiunile ei fundamentale, esenţiale.

Modelul econometric este o prezentare formalizată a realităţii economice şi ia sub forma unei ecuaţii sau al unui sistem de ecuaţii construit pe baza variabilelor statistice. Modelele econometrice mai sunt denumite şi *modele de regresie*.

### Tipuri de modele

Modelele econometrice se pot grupa în mai multe categorii, în funcţie de diferite criterii: natura dependenţei dintre variabile, numărul de factori de influenţă, forma legăturii dintre variabile şi timpul la care se referă datele din model.

### Modele de regresie deterministe şi modele de regresie probabiliste

Dependenţa dintre variabilele studiate poate fi fie deterministă, fie stochastică.

- Dependenţa deterministă sau funcţională este de tipul $y_i = f(x_i)$. De exemplu, astfel de relaţii de dependenţă se întâlnesc în cazul legilor fizicii din sistemul newtonian sau a relaţiilor funcţionale care definesc anumiţi indicatori economici (productivitatea muncii, de exemplu). În acest caz, se consideră că variabila dependentă este explicată în totalitate de variabila sau de variabilele independente care apar în model.
- Dependenţa stochastică sau probabilistă are loc în cazul în care variabila dependentă este influenţată şi de o sumă de factori care nu pot fi comensuraţi, dar sunt luaţi în considerare printr-o variabilă numită *eroare* sau *reziduu* şi este de tipul $y_i = f(x_i) + \varepsilon_i$. Componenta stochastică apare în model printr-o variabilă aleatoare notată, de regulă, cu $\varepsilon$.

### Modele de regresie simplă şi modele de regresie multiplă

Într-un model de regresie poate să apară o variabilă independentă sau mai multe variabile independente.

În cazul unei singure variabile independente, regresia se numeşte *simplă*. Modelele de regresie simplă se mai numesc şi unifactoriale şi au la bază ipoteza că printre factorii explicativi pentru o variabilă dependentă există un singur factor determinant, iar ceilalţi factori au o acţiune aleatoare sau nesemnificativă. Un exemplu de model simplu este modelul care exprimă dependenţa consumului de venituri: $C = f(V) + \varepsilon$.

Regresia se numeşte *multiplă* când există două sau mai multe variabile independente în model. Un exemplu este modelul funcţiei de producţie: $Q = f(L, K) + \varepsilon$, unde $Q$ (producţia) este output-ul, iar $L$, $K$ reprezintă input-ul ($L$ este factorul muncă şi $K$ este capitalul).

### Modele de regresie liniare şi modele de regresie neliniare

Un model de regresie $y = f(x)$ este liniar în $X$ dacă variabila independentă apare în model la puterea întâi şi nu este înmulţită sau împărţită printr-o altă variabilă, sau, cu alte cuvinte, dacă $f$ este o funcţie de gradul întâi în $X$. Un exemplu de model liniar este $Y_X = \alpha + \beta X$ sau $Y_X = \alpha + \beta_1 X_1 + \beta_2 X_2$.

Modelul de regresie este neliniar în celelalte cazuri. De exemplu, sunt neliniare următoarele modele de regresie: $Y_X = \alpha + \beta_1 X + \beta_2 X^2$, $Y_X = \alpha + \beta_1 X_1 X_2$, $Y_X = \alpha \cdot X_1^{\beta_1} \cdot X_2^{\beta_2}$

### Model de regresie static şi model de regresie dinamic

Modelul econometric în care legătura dintre variabila dependentă şi cele independente se referă la acelaşi moment de timp sau la aceeaşi perioadă de timp se numeşte *model static* sau *de moment*. Astfel de modele se construiesc, de regulă, pe baza datelor de sondaj sau a cercetărilor de moment.

Modelele econometrice dinamice sunt modele în care apare factorul timp, fie explicit, ca variabilă independentă, fie implicit, prin introducerea unui decalaj la nivelul variabilelor. Dacă în model variabila dependentă apare ca variabilă independentă cu valori decalate în timp, modelul se numeşte *autoregresiv*, iar dacă variabila eroare apare cu valori decalate, modelul se numeşte de tip *medie mobilă*.

### Variabile

În cercetarea econometrică se utilizează variabile statistice între care există relaţii de interdependenţă. Tipuri de variabile:

- *variabile dependente*, numite şi *variabile rezultative* sau *efect*, *rezultat*;
- *variabile independente*, numite şi *variabile factoriale* sau *factori de influenţă* care determină un anumit efect asupra variabilei rezultat.
- *variabilele reziduale*, numite şi *erori*, care apar în model ca sumă a tuturor influenţelor necunoscute sau variabile care nu apar explicit în model. În cercetarea econometrică, variabila eroare este o variabilă aleatoare care respectă anumite proprietăţi exprimate ca şi ipoteze clasice.

### Parametri

Parametrii modelului econometric, numiţi şi coeficienţi de regresie, sunt mărimi reale şi necunoscute care apar în model în diferite expresii alături de variabile. Parametrii fac obiectul procesului de estimare şi testare statistică.

### Estimatori

Estimatorii sunt variabile aleatoare, construite în procesul de estimare, cu distribuţii de probabilitate cunoscute şi cu proprietăţi specifice în baza cărora se realizează procesul de estimare a parametrilor modelului econometric.

Notăm parametrul cu simbolul $\theta$ şi un estimator al acestuia cu $\hat{\theta}$.

În procesul de estimare, cele mai importante proprietăţi ale estimatorilor sunt:

- *nedeplasarea* – un estimator este nedeplasat dacă media sau speranţa matematică a acestuia este egală cu parametrul. Un estimator nedeplasat verifică relaţia: $M(\hat{\theta})=\theta$. Dacă relaţia nu este respectată, atunci estimatorul este deplasat;

- *convergenţa* – un estimator este convergent dacă, pentru un eşantion cu volum suficient de mare, şirul estimatorilor converge în probabilitate către parametru. Pentru un estimator convergent are loc relaţia:

$$\lim_{n\to\infty} P(\left|\hat{\theta}_n-\theta\right|<\varphi)=0,\quad (\forall)\varphi\in(0,1)\ ;$$

- *eficienţa* – estimatorul $\hat{\theta}$ este eficient dacă are dispersia sau varianţa cea mai mică dintre toţi estimatorii posibili pentru parametru $\theta$.

## Estimaţii

Estimaţiile sunt valori ale estimatorilor calculate la nivelul unui eşantion sau al unui set de date reale observate din realitate.

## Ipoteze statistice

Ipotezele statistice sunt presupuneri cu privire la parametrii modelului econometric sau cu privire la legea de repartiţie a unei variabile din model. Există două tipuri de ipoteze:

- Ipoteza nulă, pe care, de regulă, ne propunem să o respingem în urma procesului de testare. Se notează cu $H_0$. În cazul parametrilor unui model de regresie, ipoteza se scrie astfel: $H_0:\beta_i=0$;
- Ipoteza alternativă, pe care, de regulă, ne propunem să o acceptăm în urma procesului de testare. Se notează cu $H_1$ şi este complementara ipotezei nule. $H_1:\beta_i\neq 0$.

## Test statistic

Un test statistic sau o statistică este o variabilă aleatoare cu legi de repartiţie cunoscute şi complet specificate. Este folosit în testarea statistică, care este un procedeu prin care se verifică o ipoteză statistică. La finalul acestui procedeu, pe baza unei reguli, se ia decizia de a accepta sau de a respinge ipoteza nulă.

## 1.4 Demers metodologic

Demersul cercetării în econometrie se poate rezuma sintetic la parcurgerea următoarelor etape:

- formularea problemei, pornind de la o teorie economică;
- identificarea variabilelor;
- identificarea tipului şi formei legăturii dintre variabile, după o analiză atentă a fenomenului real şi a teoriei economice;
- propunerea unuia sau a mai multor modele care explică realitatea studiată prin relaţii de dependenţă între variabile;
- estimarea parametrilor modelului sau modelelor propuse, pe baza metodelor statistice cunoscute;
- testarea modelului sau modelelor şi alegerea celui mai bun model;
- aplicarea în practică sau realizarea de predicţii pe seama modelului.

Schematic, demersul metodologic al cercetării econometrice este prezentat în figura 1.

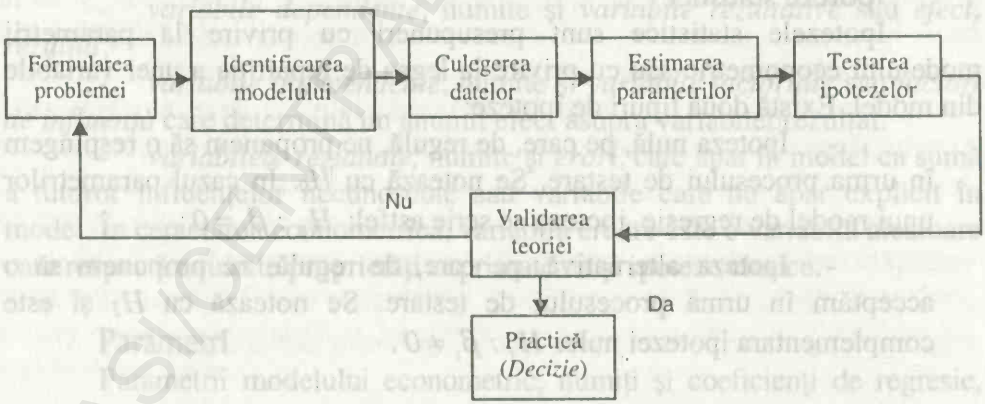



**Figura nr. 1.** *Demersul metodologic al cercetării econometrice*

Din figura 1, se poate observa că cercetarea econometrică implică o dimensiune ciclică, prin corelarea rezultatelor cu practica economică. Rezultatele cercetării econometrice pot determina, în mod direct, decizii în practică, dar pot evidenţia alte noi probleme care să fie supuse cercetării ulterioare.

BCU IASI / CENTRAL UNIVERSITY LIBRARY

## 1.5. Date statistice în cercetarea econometrică

În cercetarea econometrică, calitatea datelor reprezintă o problemă fundamentală. Calitatea poate fi analizată în funcţie de tipul datelor, de sursele de date, de metodele de culegere şi de accesibilitatea acestora. Vom analiza aceste criterii pentru trei tipuri de date utilizate mai frecvent în econometrie: seriile de timp, datele furnizate de anchetele statistice la un moment dat şi datele obţinute prin anchete de tip panel, care combină seriile de timp cu datele observate la un moment dat prin anchete.

### Seriile de timp

O serie de timp reprezintă un set de înregistrări, de date observate pentru un fenomen la diverse momente sau intervale de timp. Importanţa seriilor de timp în cercetarea economică derivă din faptul că timpul joacă un rol determinant în viaţa economică, astfel că un set de date sau de observaţii statistice depinde de cele mai multe ori de factorul timp. Astfel, fenomene economice din trecut pot influenţa semnificativ fenomene viitoare, iar această influenţă trebuie luată în considerare în analiza econometrică.

Datele statistice care formează seriile de timp sunt înregistrate pentru momente sau perioade diferite, de la înregistrări orare, până la cele zilnice, săptămânale, lunare sau anuale. Variabilele observate pot fi numerice, nominale sau ordinale. Câteva probleme pot fi evidenţiate în legătură cu datele economice sub formă de serii de timp:

- datele sunt accesibile pentru un număr restrâns de variabile, de obicei culese prin sistemul informaţional statistic oficial;
- datele sunt agregate la un nivel ridicat (în general, datele nu sunt disponibile decât la nivel naţional şi în puţine cazuri la nivel regional);
- pentru perioade lungi de timp, datele pot suferi sub aspectul comparabilităţii, datorită modificării metodologiilor de înregistrare şi de calcul.

### Datele obţinute din anchete statistice

Aceste date vizează colectivităţi statistice bine delimitate, cum ar fi, de exemplu, populaţia unei ţări, colectivitatea firmelor dintr-o regiune etc. Datele se obţin atât prin cercetări exhaustive de tip recensământ sau prin anchete parţiale, prin cercetări prin sondaj.

Principalele probleme ale datelor rezultate din anchetele prin sondaj sunt:

- reprezentativitatea (dacă provin din eşantioane aleatoare sau mixte care permit inferenţa statistică);
- gradul de omogenitate a acestora (reprezentativitatea mărimilor sintetice, îndeosebi a indicatorilor medii);
- comparabilitatea în timp şi în spaţiu.

### Datele obţinute din anchete de tip panel

Anchetele de tip panel reprezintă anchete prin sondaj care se realizează permanent sau la intervale regulate de timp pe acelaşi eşantion de unităţi sau pe eşantioane care sunt permanent modificate, însă care păstrează aceleaşi caracteristici de omogenitate.

Din anchetele panel se obţin ambele categorii de date prezentate mai sus, atât serii de timp, dacă se analizează rezultatele anchetelor pe o perioadă de timp, cât şi date de moment, dacă se consideră o anchetă realizată la un anumit moment. Prin aceste anchete sunt ameliorate o serie de probleme de calitate a datelor obţinute din anchete sau sub formă de serii de timp: reprezentativitatea, comparabilitatea, accesibilitatea etc.

## 1.6. Notaţii

În lucrare, se vor utiliza următoarele notaţii:

$Y$ – variabila dependentă;

$X_i$ – variabilele independente, $i = \overline{1,p}$, unde

$k$ – numărul de parametri din model, $k=p+1$;

$\varepsilon$ – variabila reziduală sau eroare;

$Y = f(X_i)+\varepsilon$ – modelul econometric, $i = \overline{1,p}$;

$\alpha, \beta_i$ – parametrii modelului, $i = \overline{1,p}$

(parametrul $\alpha$ este coeficientul care însoţeşte variabila degenerată $X=1$ sau variabila „constantă", parametrii $\beta_i$ sunt coeficienţii care apar în model alături de variabilele independente în diferite expresii);

$\hat{\alpha}, \hat{\beta}_i$ – estimatorii parametrilor modelului $i = \overline{1,p}$;

$a, b_i$ – estimaţii ale parametrilor modelului, $i = \overline{1,p}$;

$n$ – volumul eşantionului;

$N$ – volumul populaţie totale.

# Capitolul 2

# MODELE DE REGRESIE SIMPLĂ

1. Modelul liniar simplu
   1.1. Formularea problemei. Exemple din economie
   1.2. Model şi ipoteze
   1.3. Estimarea parametrilor modelului
   1.4. Testarea parametrilor modelului
   1.5. Coeficienţi de corelaţie
   1.6. Regresia liniară simplă în SPSS
2. Modelul neliniar simplu
   2.1. Formularea problemei şi exemple din economie
   2.2. Modele liniarizabile
   2.3. Modele polinomiale

În acest capitol, sunt prezentate cele mai simple modele econometrice – modelele care conţin o variabilă dependentă şi o variabilă independentă între care poate exista o legătură de tip liniar sau una neliniară.

## 2.1. Modelul liniar simplu

Modelul liniar simplu este modelul în care relaţia dintre cele două variabile este descrisă cu ajutorul unei ecuaţii de gradul întâi. Legătura liniară dintre două variabile presupune că între acestea există o relaţie de

proporţionalitate, adică variaţia unei variabile se realizează într-o relaţie de proporţionalitate cu variaţia celeilalte variabile.

### 2.1.1. Formularea problemei. Exemple din economie

Modelul liniar simplu este cea mai simplă schemă explicativă a legăturii dintre două variabile.

În economie, există situaţii în care un rezultat sau un fenomen poate fi explicat într-o proporţie ridicată doar de influenţa unui singur factor. Acest factor apare în modelul econometric drept variabilă independentă, iar restul influenţelor este preluat de variabila reziduală.

#### Exemple din economie de modele liniare simple

*Funcţia de consum*

Cererea sau consumul populaţiei pentru o anumită categorie de mărfuri este o funcţie de venit:

$$c_i = \alpha + \beta v_i + \varepsilon_i .$$

În modelul de mai sus, parametrul $\beta$ este, de regulă, pozitiv şi arată de câte ori creşte, în medie, consumul unui anumit produs ($c_i$) la o creştere cu o unitate a venitului ($v_i$).

*Legea cererii*

Cererea populaţiei pentru o anumită categorie de mărfuri este în funcţie de preţul acestor produse:

$$c_i = \alpha + \beta p_i + \varepsilon_i ,$$

În model, parametrul $\beta$ este, de regulă, negativ şi arată cu cât scade, în medie, cererea ($c_i$) la o creştere a preţului ($p_i$) cu o unitate.

### 2.1.2. Model şi ipoteze

Modelul de regresie liniară simplă are următoarea expresie formală:

$$Y = \alpha + \beta X + \varepsilon$$

Relaţia de mai sus se numeşte **ecuaţie de regresie** şi este compus din: funcţia liniară $Y_X = \alpha + \beta X$ şi eroarea $\varepsilon$.

Variabilele din ecuație sunt:
*Y – variabila dependentă, aleatoare*;
*X – variabila independentă, nonaleatoare*;
*ε – variabila aleatoare eroare sau reziduu.*

### Componentele modelului

Modelul prezentat mai sus include două componente: una deterministă și una stochastică.

Componenta deterministă este reprezentată de variabila independentă (*X*), care este o variabilă observabilă din punct de vedere statistic. Pentru fiecare factor sau variabilă independentă din model este important să se precizeze condițiile de calitate a observării statistice, modalitatea de culegere a datelor, momentul înregistrării, prelucrările preliminare, erorile de înregistrare etc.

Componenta aleatoare prezentă în modelul liniar de regresie se mai numește variabilă eroare ($\varepsilon$) sau reziduu și se referă, în sinteză, la următoarele probleme[1] care însoțesc procesul de modelare: problema de specificare a variabilelor din model, problemele generate de natura fenomenului studiat, problemele legate de erorile de măsurare.

### Problema de specificare a variabilelor din model (*vagueness of theory*)

Din perspectiva specificării modelului, variabila eroare substituie toate celelalte variabile care nu au fost incluse în model din anumite motive: lipsa datelor, dificultatea unui model cu mai multe variabile, lipsa unei teorii economice etc. De exemplu, în cazul studiului producției sau output-ului, pe lângă factorii economici considerați clasici (muncă, capital) există o serie de factori noi, decisivi pentru condițiile economice actuale (informația, inovarea etc.), factori care fie nu pot fi prezenți în model din lipsa datelor, fie sunt neglijați.

### Problemele generate de natura fenomenului studiat (*intrinsec randomness in human behavior*)

Variabila aleatoare include acele influențe care nu pot fi explicitate și explicate prin variabile în model. Comportamentul economic este

[1] D.N. Gujarati, *Basic econometrics*, McGraw-Hill, 1995, p. 39.

determinat nu numai de relaţii de cauzalitate în sens determinist. La baza comportamentului economic stă comportamentul uman în general, bazat pe judecăţi de valoare obiective, dar şi pe criterii subiective, care nu pot fi prevăzute.

### Problemele legate de erorile de măsurare şi de eşantionare (*methodological problems*)

Fenomenele economice sunt realităţi care nu pot fi cunoscute în mod direct, ci prin intermediul măsurătorilor, prin observare statistică în timpul şi la locul unde acestea se manifestă. Modelarea econometrică utilizează date statistice culese în anumite condiţii de calitate (care implică erori de observare) şi, de cele mai multe ori, sunt date obţinute prin metode de sondaj, adică sunt date estimate, care conţin şi erori de reprezentativitate specifice fiecărei metode de sondaj. Componenta aleatoare din model înglobează aceste erori determinate de metoda statistică.

Modelarea econometrică presupune anumite condiţii sau ipoteze asupra celor două componente ale modelului, ipoteze care vor fi prezentate în continuare.

### Ipotezele modelului de regresie

Cele mai importante ipoteze ale modelului de regresie sunt:

- normalitatea erorilor la nivelul distribuţiilor condiţionate de tipul $Y_i|X = x_i : \varepsilon_i \sim N(0,\sigma^2)$, adică variabila reziduală urmează o lege de repartiţie normală de medie zero şi varianţă $\sigma^2$;
- homoscedasticitate: $V(\varepsilon_i) = M(\varepsilon_i^2) = \sigma^2$, adică varianţa erorii este constantă la nivelul distribuţiilor condiţionate de tipul $Y_i|X = x_i$;
- necorelarea erorilor: $cov(\varepsilon_i, \varepsilon_j) = 0$, adică erorile nu se influenţează reciproc;
- lipsa corelaţiei dintre variabila independentă şi variabila eroare: $cov(\varepsilon, X) = 0$.

Parametrii ecuaţiei de regresie sunt:

$\alpha$ – *ordonata la origine*[2], care arată valoarea variabilei $Y$ când $X = 0$;

[2] Simbolul $\alpha$ utilizat aici nu trebuie confundat cu simbolul ce desemnează pragul de semnificaţie al unui test.

*$\beta$ – panta dreptei.*

În ecuația de regresie, $\alpha$ și $\beta$ sunt valori necunoscute. Semnul parametrului de regresie $\beta$ indică direcția legăturii dintre cele două variabile corelate:

$\beta > 0$ legătură directă (pozitivă);

$\beta = 0$ nu există legătură;

$\beta < 0$ legătură inversă (negativă).

Parametrul de regresie $\beta$ arată gradul de dependență dintre variabile, respectiv cu cât crește sau scade în medie $Y$ la o creștere sau la o scădere a variabilei $X$ cu o unitate:

$$\beta = \frac{dY}{dX}.$$

Cu alte cuvinte, $\beta$ arată variația absolută a variabilei dependente la o variație absolută de o unitate a variabilei independente.

Conform teoriei probabilităților și a statisticii matematice, regresia este o medie condiționată definită pe o distribuție bi- sau multi-dimensională. În cazul unei legături dintre două variabile, regresia este definită prin aplicația:

$$M(Y/X = x_i) = f(x_i) \text{ sau } M(Y/X) = f(x)$$

Pentru cazul liniar, regresia sau media condiționată este o funcție liniară de forma:

$M(Y/X) = \alpha + \beta X$, unde $\alpha$, $\beta$ sunt parametrii modelului, iar $X$ reprezintă variabila independentă.

În consecință, modelul statistic de regresie liniară este o funcție de forma:

$$y_i = M(Y/X = x_i) = \alpha + \beta x_i.$$

Modelul de regresie liniară simplă este dat de ecuația:

$$Y = M(Y/X) + \varepsilon = \alpha + \beta X + \varepsilon$$

BCU IASI CENTRALA UNIVERSITARA LIBRARY

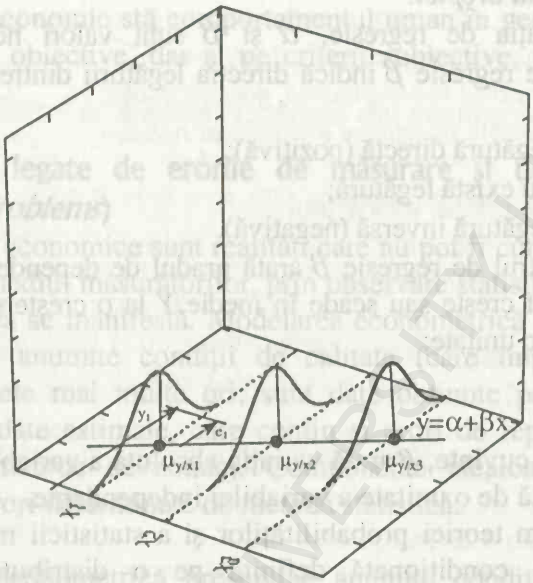



**Figura nr. 1.** *Legi de repartiţie condiţionate* $Y_i / X = x_i$

## 2.1.3. Estimarea parametrilor modelului

În practică, determinarea parametrilor la nivelul populaţiei totale nu este posibil de realizat, fapt care impune estimarea parametrilor.

Parametrii modelului de regresie se estimează pe baza estimatorilor $\hat{\alpha}$ şi $\hat{\beta}$. Folosind date înregistrate asupra unui eşantion de $n$ perechi de observaţii ale variabilelor $X$ şi $Y$, se calculează estimaţiile $a$ şi $b$ ale parametrilor $\alpha$ şi $\beta$.

### A. Estimarea punctuală a parametrilor ecuaţiei de regresie

Ecuaţia estimată a modelului de regresie este de forma:

$Y_x = a + bX$

BCU IASI / CENTRAL UNIVERSITY LIBRARY

În relația de mai sus, *a* și *b* reprezintă *valori ale estimatorilor* $\hat{\alpha}$ *și* $\hat{\beta}$, respectiv *estimații ale parametrilor* $\alpha$ *și* $\beta$, calculate la nivelul unui eșantion de volum *n*.

Se consideră acei estimatori $\hat{\alpha}$ și $\hat{\beta}$, ai parametrilor $\alpha$ și $\beta$, pentru care valoarea $e_i$ a variabilei eroare $\varepsilon$, adică distanța dintre o valoare observată, $y_i$, și o valoare estimată a ecuației de regresie la nivelul eșantionului, $y_{x_i} = a + bx_i$, să fie cât mai mică. Adică:

$$e_i = y_i - y_{x_i} = y_i - (a + bx_i)$$

să fie minimă pentru orice valoare „*i*" a variabilei *X*.

Estimarea parametrilor se poate face pornind de la diferite criterii, cum ar fi:

1. $\min_{1 \le i \le n} \{|e_i|\}$
2. $\sum |e_i| : minim$
3. $\sum (e_i)^2 : minim$.

De regulă, în practică se utilizează ultimul criteriu, care definește **metoda celor mai mici pătrate (MCMMP).**

Aplicarea MCMMP presupune minimizarea expresiei:

$$S = \sum e_i^2 = \sum (y_i - y_{x_i})^2 = min\ im$$

Înlocuind valoarea $y_{x_i}$, obținem:

$$S = \sum (y_i - a - bx_i)^2 = minim$$

Rezolvarea problemei de minim impune două condiții:

1. anularea derivatelor parțiale de ordinul întâi ale lui *S* în raport cu *a* și *b;*
2. matricea derivatelor parțiale de ordinul doi să fie definită pozitiv.

**1. Derivatele parțiale de ordinul întâi:**

$$\frac{\partial S}{\partial a} = 2\sum (y_i - a - bx_i)(-1) = 0$$

$$\frac{\partial S}{\partial b} = 2\sum (y_i - a - bx_i)(-x_i) = 0, \quad i = \overline{1,n}$$

din care obținem un sistem de ecuații normale sub forma:

$$na + b\sum x_i = \sum y_i, \quad i = \overline{1,n}$$
$$a\sum x_i + b\sum x_i^2 = \sum x_i y_i$$

**2. Derivatele parțiale de ordinul doi:**

$$\frac{\partial^2 S}{\partial a^2} = 2n, \quad \frac{\partial^2 S}{\partial a \partial b} = 2\sum x_i, \quad \frac{\partial^2 S}{\partial b^2} = 2\sum x_i^2$$

Matricea derivatelor parțiale de ordinul doi:

$$\begin{pmatrix} n & \sum x_i \\ \sum x_i & \sum x_i^2 \end{pmatrix},$$

este pozitiv definită, deoarece $n\sum x_i^2 - (\sum x_i)^2 = n\sigma^2 > 0$

Prin rezolvarea sistemului de ecuații normale, utilizând una dintre metodele cunoscute (metoda determinanților, metoda Doolittle etc.), se obțin estimațiile $a$ şi $b$.

Aplicând metoda determinanților, se obțin pentru $a$ şi $b$ următoarele relații de calcul:

$$b = \frac{\Delta b}{\Delta} = \frac{n\sum x_i y_i - \sum x_i \sum y_i}{n\sum x_i^2 - (\sum x_i)^2}, \quad i = \overline{1,n}$$

$$a = \frac{\Delta a}{\Delta} = \frac{\sum y_i \sum x_i^2 - \sum x_i \sum x_i y_i}{n\sum x_i^2 - (\sum x_i)^2} \quad \text{sau}$$

$$a = \bar{y} - b\bar{x}$$

În relația de mai sus, $a$ şi $b$ reprezintă valori de sondaj, estimații ale parametrilor $\alpha$ şi $\beta$, calculate la nivelul unui eşantion prin aplicarea metodei celor mai mici pătrate.

## B. Estimarea prin interval de încredere

Estimarea prin interval de încredere se bazează pe distribuțiile de selecție ale estimatorilor $\hat{\alpha}$ şi $\hat{\beta}$ ai parametrilor $\alpha$ şi $\beta$.

Pentru modelul liniar simplu, se poate demonstra că *estimatorii parametrilor obținuți prin MCMMP urmează o lege de distribuție normală* şi sunt *nedeplasați, convergenți* şi *eficienți*.

BCU IASI / CENTRAL UNIVERSITY LIBRARY

Au loc relațiile:

$\hat{\alpha} \sim N(\alpha, \sigma_{\hat{\alpha}}^2)$; unde

$$M(\hat{\alpha}) = \alpha;\ V(\hat{\alpha}) = \sigma_{\hat{\alpha}}^2;\ \sigma_{\hat{\alpha}}^2 = \frac{\sum_i x_i^2}{n\sum_i (x_i - \bar{x})^2}\sigma^2,\ \sigma^2 = V(\varepsilon)$$

$\hat{\beta} \sim N(\beta, \sigma_{\hat{\beta}}^2)$; unde

$$M(\hat{\beta}) = \beta;\ V(\hat{\beta}) = \sigma_{\hat{\beta}}^2;\ \sigma_{\hat{\beta}}^2 = \frac{\sigma^2}{\sum_i (x_i - \bar{x})^2},$$

Se observă că dispersia estimatorilor parametrilor depinde de dispersia erorilor, care este un parametru, deci nu este cunoscută. În practică, se utilizează un estimator al dispersiei erorilor de forma:

$$\hat{\sigma}_{\hat{\varepsilon}}^2 = \frac{\sum_i \hat{\varepsilon}_i^2}{n-2} = \frac{\sum_i (y_i - \hat{\alpha} - \hat{\beta}x_i)^2}{n-2}$$

Pentru parametrii din relațiile de mai sus, sunt valabile următoarele estimații:

- pentru varianța erorilor $\sigma^2$:

$$s_{\hat{\varepsilon}}^2 = \frac{\sum_i e_i^2}{n-2} = \frac{\sum_i (y_i - a - bx_i)^2}{n-2}$$

- pentru varianța estimatorului $\hat{\alpha}$:

$$s_{\hat{\alpha}}^2 = \frac{\sum_i x_i^2}{n\sum_i (x_i - \bar{x})^2} s_{\hat{\varepsilon}}^2$$

- pentru varianța estimatorului $\hat{\beta}$:

$$s_{\hat{\beta}}^2 = \frac{s_{\hat{\varepsilon}}^2}{\sum_i (x_i - \bar{x})^2}$$

Intervalul de încredere pentru coeficientul de regresie $\beta$, estimat pentru un eșantion observat, este definit de relația:

$$b \pm t_{\alpha/2} \cdot s_{\hat{\beta}}$$

BCU IASI CENTRAL UNIVERSITY LIBRARY

Analog, pentru parametrul $\alpha$, se determină intervalul:

$a \pm t_{\alpha/2} \cdot s_{\hat{a}}$

**Exemplu.** Considerând datele din tabelul 1, se cere să se estimeze parametrii modelului de regresie corespunzător.

**Tabelul nr. 1.** *Date convenționale*

| Îngrășăminte (u.c.) *(X)* | Prod. medie de grâu/ha (u. p.) *(Y)* |
|---|---|
| 1,0 | 10 |
| 2,0 | 15 |
| 3,0 | 20 |
| 4,0 | 30 |
| 5,0 | 40 |

Corelograma din figura 2 evidențiază o legătură directă, liniară între cantitatea de îngrășăminte și producția obținută la hectar. Prin urmare, pentru cazul considerat, estimarea producției în funcție de cantitatea de îngrășăminte se efectuează cu ajutorul ecuației de regresie liniară:

$Y_x = a + bX$ .

Elementele de calcul necesare pentru determinarea parametrilor ecuației de regresie sunt sistematizate în tabelul 2.

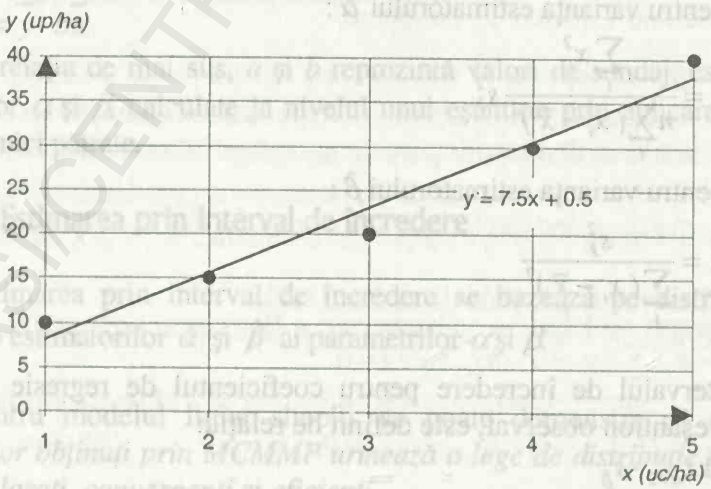



**Figura nr. 2.** *Legătura dintre cantitatea de îngrășăminte și producția medie de grâu / ha*

BCU IASI CENTRAL UNIVERSITARA

Ecuația estimată este:

$$Y_x = a + bX = 0{,}5 + 7{,}5X$$

Estimația parametrului de regresie $\beta$, luând valoare pozitivă, arată că legătura între variabilele $X$ și $Y$ este directă. De asemenea, aceasta scoate în evidență gradul de dependență dintre variabile, și anume la o creștere a lui $X$ cu o unitate (uc/ha), $Y$ crește în medie cu *7,5* unități.

**Tabelul nr. 2.** *Elemente de calcul*

| $x_i$ | $y_i$ | $x_i^2$ | $x_i y_i$ | $y_i^2$ | $y_{x_i}$ | $x_i - \bar{x}$ | $(x_i - \bar{x})^2$ |
|---|---|---|---|---|---|---|---|
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
| 1 | 10 | 1 | 10 | 100 | 8,0 | -2 | 4 |
| 2 | 15 | 4 | 30 | 225 | 15,5 | -1 | 1 |
| 3 | 20 | 9 | 60 | 400 | 23,0 | 0 | 0 |
| 4 | 30 | 16 | 120 | 900 | 30,5 | 1 | 1 |
| 5 | 40 | 25 | 200 | 1600 | 38,0 | 2 | 4 |
| 15 | 115 | 55 | 420 | 3225 | 115,0 | - | 10 |

## 2.1.4. Testarea parametrilor modelului

În testarea ipotezelor parametrilor modelului de regresie, se utilizează următorul demers:

### Formularea ipotezelor

Testarea semnificației coeficientului de regresie $\beta$ pleacă de la formularea următoarelor ipoteze:

$$H_0 : \beta = 0$$

$$H_1 : \beta \neq 0$$

Ipoteza $H_0$ indică lipsa unei legături dintre variabile sau că cele două variabile sunt independente, iar ipoteza alternativă indică existența unei legături liniare între cele două variabile. Dacă respingem ipoteza $H_0$, cu un prag de semnificație $\alpha$ ales, atunci legătura dintre cele două variabile $X$ și $Y$ este semnificativă. În practica economică, se consideră, de regulă, un

$\alpha = 0,05$, adică se consideră un risc de 5% de a respinge ipoteza $H_0$ atunci când aceasta ar fi adevărată.

### Testul statistic

Pentru testarea semnificației coeficientului de regresie $\beta$, se folosește *statistica Student* definită de raportul:

$$t = \frac{\hat{\beta} - \beta}{\hat{\sigma}_{\hat{\beta}}}$$

În ipoteza $H_0$, statistica $t = \frac{\hat{\beta} - \beta}{\hat{\sigma}_{\hat{\beta}}}$ devine:

$$t = \frac{\hat{\beta} - 0}{\hat{\sigma}_{\hat{\beta}}} = \frac{\hat{\beta}}{\hat{\sigma}_{\hat{\beta}}},$$

care este o statistică ce urmează o lege de repartiție *Student* de *(n-2)* grade de libertate.

La nivelul unui eșantion observat, raportul *t* se scrie:

$$t_{calc.} = \frac{b}{s_{\hat{\beta}}}$$

Estimația variației estimatorului parametrului de regresie $\beta$, la nivelul unui eșantion observat, se calculează după relația:

$$s^2_{\hat{\beta}} = \frac{s^2_{\hat{\varepsilon}}}{\sum_i (x_i - \bar{x})^2}$$

Pentru un prag de semnificație $\alpha$, se citește din tabelul Student o valoare teoretică a testului $t_{\alpha/2;n-2}$ care va fi comparată cu valoarea calculată la nivelul eșantionului observat. Se utilizează un risc $\alpha/2$ pentru determinarea valorii teoretice deoarece s-a construit un test bilateral. Distribuția Student este simetrică, iar suprafața de respingere ($\alpha$) este împărțită în două părți egale ($\alpha/2$) – vezi figura 3.

BCU IAȘI CENTRAL UNIVERSITY LIBRARY

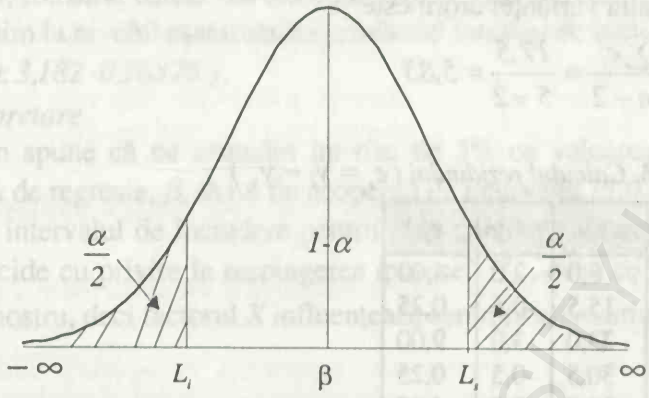



**Figura nr. 3.** *Distribuția de selecție a estimatorului $\hat{\beta}$ și intervalul de încredere*

### Semnificația testului și regula de decizie

În practica statistică, este mult mai simplu ca regula de decizie cu privire la acceptarea sau respingerea ipotezei nule să se bazeze pe compararea probabilității cu care se obține valoarea calculată a testului și a pragului de semnificație. Această probabilitate calculată a primit numele de *p-value* (*Sig*) și este probabilitatea exactă cu care se comite eroarea de speța întâi în procesul de estimare sau probabilitatea cu care se acceptă ipoteza nulă.

*P-value* se calculează ca probabilitatea asociată valorii calculate a testului, pe baza funcției de repartiție, utilizând tabelele probabiliste. De exemplu, dacă pentru un test Student s-a obținut o valoare calculată egală cu 3,49, pentru 40 grade de libertate, *Sig t* are valoarea 0.0015.

Astfel, în numeroase lucrări de specialitate, precum și în principalele pachete de programe informatice de statistică, regula de decizie este următoarea:

- dacă semnificația testului (*sig*) sau *p-value* este mai mică decât pragul de semnificație $\alpha$, atunci se respinge ipoteza nulă;
- invers, dacă probabilitatea calculată este mai mare decât cea stabilită, atunci se acceptă ipoteza nulă.

**Calculul estimației variației erorii.** Pe baza exemplului de mai sus, s-au calculat: $b = 7{,}5$ și $\sum(x_i - \bar{x})^2 = 10$ (coloanele 7 și 8 din tabelul 2). Valorile $s_{\hat{\beta}}$ și $s_{\hat{\varepsilon}}$ se pot calcula pe baza elementelor de calcul din tabelul 3.

Estimația varianței erorii este:

$$s_{\hat{\varepsilon}}^2 = \frac{\sum e_i^2}{n-2} = \frac{17,5}{5-2} = 5,83.$$

**Tabelul nr. 3.** *Calculul reziduului* ( $e_i = y_i - y_{x_i}$ )

| $y_i$ | $y_x$ | $e_i$ | $e_i^2$ |
|---|---|---|---|
| 10 | 8,0 | 2,0 | 4,00 |
| 15 | 15,5 | -0,5 | 0,25 |
| 20 | 23,0 | -3,0 | 9,00 |
| 30 | 30,5 | -0,5 | 0,25 |
| 40 | 38,0 | 2,0 | 4,00 |
| 115 | 115 | 0,0 | 17,50 |

Estimația varianței estimatorului $\hat{\beta}$:

$$s_{\hat{\beta}}^2 = \frac{s_{\hat{\varepsilon}}^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2} = \frac{5,83}{10} = 0,583\,;\; s_{\hat{\beta}} = 0,76376$$

Calculul raportului Student:

$$t_{calc} = \frac{b}{s_{\hat{\beta}}} = \frac{7,5}{0,76376} = 9,8198$$

**Regula de decizie.** Pentru un risc $\alpha = 0,05$, dacă $|t_{calc}| > t_{\alpha/2;n-2}$ se respinge ipoteza $H_0$, adică coeficientul de regresie $\beta$ este considerat semnificativ diferit de *0*.

În exemplul considerat, $t_{calc} = 9,82$, iar din tabelul Student citim, pentru $\alpha/2 = 0,025$ şi *n-2=3*, valoarea $t_{0,025;3} = 3,182$. Ca urmare, pentru $t_{calc.} > t_{0,025;3}$, coeficientul de regresie $\beta$ este semnificativ diferit de *0*, adică variabila *X* explică variabila *Y*.

**Determinarea intervalului de încredere.** Intervalul de încredere pentru coeficientul de regresie $\beta$ este prezentat în figura 3 şi este definit de relația:

$$b \pm t_{\alpha/2} \cdot s_{\hat{\beta}}.$$

BCU IASI / CENTRAL UNIVERSITY LIBRARY

Astfel, folosind datele din exemplul considerat anterior, pentru un risc $\alpha = 0,05$, găsim la nivelul eşantionului următorul interval de încredere:

$(7,5 \pm 3,182 \cdot 0,76376)$.

*Interpretare*

Putem spune că ne asumăm un risc de *5%* ca valoarea adevărată a coeficientului de regresie, $\beta$, să nu fie acoperită de intervalul *[5,07; 9,93]*.

Dacă intervalul de încredere pentru $\beta$ ar conţine valoarea *0* atunci nu s-ar putea decide cu privire la respingerea ipotezei $H_0$, ceea ce nu este cazul în exemplul nostru, deci factorul *X* influenţează semnificativ variabila *Y*.

## 2.1.5. Coeficienţi de corelaţie

Alături de stabilirea liniei de regresie, care exprimă modelul legăturii între variabile, este necesar să se măsoare şi intensitatea legăturii. Aceasta arată gradul de concentrare sau de împrăştiere a valorilor $y_i$ în jurul liniei de regresie $y_x$.

Intensitatea legăturii se poate măsura cu ajutorul **coeficientului de corelaţie** şi a **raportului de corelaţie**. O apreciere aproximativă a intensităţii legăturii dintre variabile se poate obţine prin aplicarea **metodelor neparametrice**.

### Coeficientul de corelaţie

Se foloseşte pentru măsurarea legăturii în cazul unei regresii liniare simple.

Pentru două variabile aleatoare *X* şi *Y*, la nivelul unei populaţii de volum *N*, coeficientul de corelaţie teoretic se notează cu $\rho(X,Y)$ şi este definit de relaţia:

$$\rho(X,Y) = \frac{cov(X,Y)}{\sigma_x \cdot \sigma_y} = \frac{\sum_i (x_i - \mu_X)(y_i - \mu_Y)}{N \cdot \sigma_x \cdot \sigma_y}, \quad i = \overline{1,N}$$

în care:

- $cov(X,Y)$ – covarianţa;
- $x_i, y_i, \mu_X, \mu_Y$ – valorile variabilelor corelate şi nivelul mediu al acestora;
- $N$ - numărul perechilor de valori $(x_i, y_i), i = \overline{1,N}$ ;
- $\sigma_x$ , $\sigma_y$ – abaterea medie pătratică pentru *X*, respectiv *Y*.

Comparând relația de determinare a coeficientului de regresie cu cea a coeficientului de corelație se constată că între acești indicatori există următoarea legătură:

$$\rho(X,Y)=\beta\cdot\frac{\sigma_x}{\sigma_y}$$

din care rezultă că semnul coeficientului de corelație coincide cu semnul coeficientului de regresie, deoarece $\sigma_x,\ \sigma_y \geq 0$.

Valoarea coeficientului de corelație este cuprinsă între *-1* și *+1*:

$-1 \leq \rho \leq +1$

Cele două valori extreme reprezintă o *legătură liniară perfectă* (funcțională) între cele două variabile, „pozitivă", respectiv, „negativă". *Valoarea 0 semnifică absența legăturii* între cele două variabile.

Coeficientul de corelație este un parametru care trebuie estimat.

## Estimarea coeficientului de corelație

Un estimator pentru $\rho$ este $\hat{\rho}$, care are ca valori posibile coeficienții de corelație empirici, determinați la nivelul eșantioanelor posibil de extras printr-o metodă de sondaj.

În acest sens, la nivelul unui eșantion de volum *n*, se determină *coeficientul de corelație empiric* propus de K. Pearson:

$$r=\frac{cov(x,y)}{s_x\cdot s_y}=\frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{n\cdot s_x\cdot s_y},$$ care reprezintă o estimație pentru parametrul $\rho$.

Dezvoltând relația de mai sus, se obține o *formulă a coeficientului de corelație empiric*, bazată pe elementele calculate deja cu ocazia determinării coeficientului de regresie:

$$r=\frac{n\sum x_i y_i-\sum x_i\sum y_i}{\sqrt{[n\sum x_i^2-(\sum x_i)^2][n\sum y_i^2-(\sum y_i)^2]}},\ i=\overline{1,n}$$

Folosind datele din tabelul 1, intensitatea legăturii dintre cantitatea de îngrășăminte și producția medie de grâu la hectar se determină, pe baza relației de mai sus, astfel:

$$r = \frac{5 \cdot 420 - 15 \cdot 115}{\sqrt{[5 \cdot 55 - (15)^2][5 \cdot 3225 - (115)^2]}} = \frac{375}{380,79} = 0,98479$$

Valoarea obţinută este foarte apropiată de *+1*, deci între cele două variabile există o legătură directă foarte strânsă.

## Testarea coeficientului de corelaţie

Testarea coeficientului de corelaţie se face în *scopul de a verifica dacă variabila factorială considerată (X) influenţează semnificativ variaţia variabilei rezultative (Y).*

Se pleacă de la emiterea ipotezei $H_0$, considerându-se pentru variaţia lui $X$ şi $Y$ un $\rho = 0$, adică lipsa unei legături între variabile.

Ipoteza nulă $H_0 : \rho = 0$

Ipoteza alternativă: $H_1 : \rho \neq 0$

Testarea ipotezei $H_0$ se face cu ajutorul *testului t.*

### Testul t

Pentru verificarea semnificaţiei coeficientului de corelaţie simplă, în condiţiile acceptării ipotezei nule, se utilizează statistica:

$$t = \frac{\hat{\rho}}{\hat{\sigma}_{\hat{\rho}}} = \frac{\hat{\rho}\sqrt{n-2}}{\sqrt{1-\hat{\rho}^2}}.$$

- $t$ este o statistică Student cu *(n-2)* grade de libertate,
- $\hat{\sigma}_{\hat{\rho}}$ este estimatorul abaterii medii pătratice a lui $\hat{\rho}$ (estimatorul lui $\rho$): $\hat{\sigma}_{\hat{\rho}} = \sqrt{\frac{1-\hat{\rho}^2}{n-2}}$.

La nivelul unui eşantion observat, se obţin relaţiile:

$$t_{calc.} = \frac{r}{s_{\hat{\rho}}} = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}, \quad s_{\hat{\rho}} = \sqrt{\frac{1-r^2}{n-2}}, \text{ unde}$$

- $r$ reprezintă coeficientul de corelaţie simplă la nivelul eşantionului,
- $n$ este numărul perechilor de valori $x$ şi $y$.

Valoarea calculată a lui $t$ se compară cu valoarea teoretică obţinută din tabelul Student, pentru $n$-2 grade de libertate şi pentru nivelul de semnificaţie

stabilit. Dacă $|t_{calc.}| > t_{tab.}$ *atunci se respinge* $H_0$ *şi se trage concluzia că între variabilele cercetate există o legătură semnificativă, deci coeficientul de corelaţie este semnificativ statistic.*

Considerând legătura dintre cantitatea de îngrăşăminte şi producţia medie de grâu la hectar, prezentată prin datele din tabelul 1, cu un număr de *5* valori *x* şi *y*, pentru care s-a găsit coeficientul de corelaţie $r = 0,98$, se calculează testul *t* astfel:

$$t_{calc.} = \frac{0,98\sqrt{5-2}}{\sqrt{1-0,98^2}} = 8,53$$

În tabelul Student, la $n - 2 = 3$ grade de libertate şi pentru un nivel de semnificaţie $\alpha = 0,01$, găsim valoarea $t_{tab.} = 5,841$.

Comparând $t_{calc.}$ cu $t_{tab.}$ se observă că: $t_{calc.} = 8,53 > t_{tab.} = 5,841$, prin urmare se respinge ipoteza nulă şi deci coeficientul de corelaţie are valoare semnificativ diferită de zero. Rezultatul se garantează cu o probabilitate de 0,99.

## Raportul de corelaţie

Este un indicator al intensităţii legăturii ce poate fi aplicat atât în cazul regresiei liniare, cât şi al celei neliniare simple sau multiple.

Raportul de corelaţie este notat cu $\eta$ şi se defineşte prin relaţia:

$$\eta = \sqrt{\frac{\sigma^2_{y_x}}{\sigma^2_y}}$$

sau $$\eta = \sqrt{1 - \frac{\sigma^2_{y/y_x}}{\sigma^2_y}}$$

în care:

$\sigma^2_y = \frac{\sum(y_i - \mu_Y)^2}{N}$, reprezintă varianţa generală, respectiv varianţa variabilei *Y* în raport cu media tuturor valorilor;

$\sigma^2_{y_x} = \frac{\sum(y_{x_i} - \mu_Y)^2}{N}$, varianţa valorilor teoretice faţă de media lor (varianţa sub influenţa factorilor esenţiali);

$\sigma^2_{y/y_x} = \frac{\sum(y_i - y_{x_i})^2}{N}$, varianța valorilor reale față de valorile teoretice ale variabilei (varianța reziduală).

Varianța generală este egală cu suma celorlalte două varianțe componente:

$$\sigma^2_y = \sigma^2_{y_x} + \sigma^2_{y/y_x},$$

Relația de mai sus se poate scrie în funcție de componentele variației: variația totală ($V_T$) este suma variației explicate ($V_E$) și a variației reziduale ($V_R$), adică: $V_T = V_E + V_R$, unde

$$V_T = \sum(y_i - \mu_Y)^2.$$

Descompunerea pe componente a variației unei variabile $Y$ dintr-un model de regresie și relațiile dintre ele sunt reprezentate grafic în diagrama 4.

Analiza variației pe componente ne arată că *raportul de corelație măsoară ponderea variației explicată prin linia de regresie în variația totală.* Ca urmare, valoarea raportului de corelație este un număr cuprins în intervalul [0,1]. Rezultă proprietatea: $0 \le \eta \le 1$.

*Valoarea, 1 pentru raportul de corelație, arată existența unei legături funcționale*, respectiv, cazul când variația variabilei $Y$ depinde numai de variația variabilei $X$, varianța reziduală fiind egală cu zero.

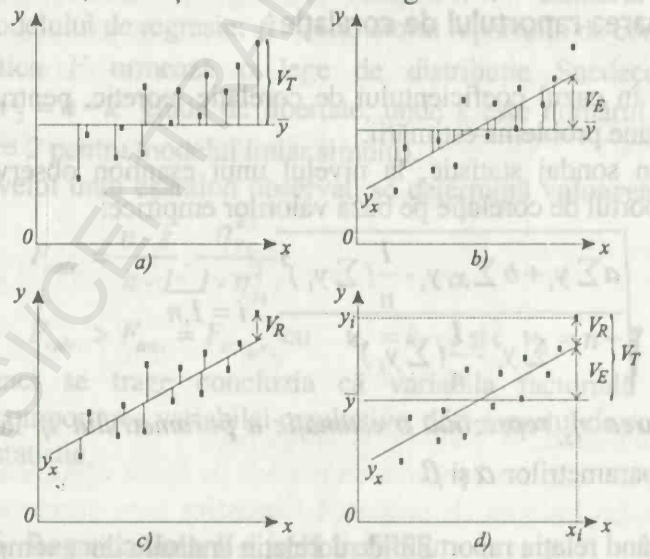



**Figura nr. 4.** *Descompunerea pe componente a dispersiei unei variabile Y, într-un model de regresie: a) varianța totală; b) varianța explicată; c) varianța reziduală; d) relația între componente.*

BCU IASI CENTRALA UNIVERSITY LIBRARY

### Coeficientul sau raportul de determinaţie

Valoarea la pătrat a raportului de corelaţie reprezintă **raportul de determinaţie:**

$$\eta^2 = \frac{\sigma^2_{y_x}}{\sigma^2_y}$$

şi *arată ponderea influenţei factorului X asupra variaţiei variabilei Y.* Acest indicator se exprimă, de regulă, în procente pentru a facilita interpretarea rezultatelor.

Prin explicitarea celor două varianţe $\sigma^2_{y/yx}$ şi $\sigma^2_y$ din raportul de corelaţie şi efectuarea unor transformări elementare se ajunge la calculul acestuia pe baza valorilor parametrilor ecuaţiei de regresie din modelul admis.

În cazul regresiei liniare ($Y = \alpha + \beta X + \varepsilon$), raportul de corelaţie devine:

$$\eta = \sqrt{\frac{\alpha \sum y_i + \beta \sum x_i y_i - \frac{1}{N}(\sum y_i)^2}{\sum y_i^2 - \frac{1}{N}(\sum y_i)^2}}, \quad i = \overline{1,N}$$

### Estimarea raportului de corelaţie

Ca şi în cazul coeficientului de corelaţie teoretic, pentru raportul de corelaţie se pune problema estimării.

Într-un sondaj statistic, la nivelul unui eşantion observat, se poate determina raportul de corelaţie pe baza valorilor empirice:

$$\eta_{y_x} = \sqrt{\frac{a \sum y_i + b \sum x_i y_i - \frac{1}{n}(\sum y_i)^2}{\sum y_i^2 - \frac{1}{n}(\sum y_i)^2}}, \quad i = \overline{1,n}$$

*Valoarea $\eta_{y_x}$ reprezintă o estimaţie a parametrului $\eta$,* iar *a* şi *b* sunt estimaţii ale parametrilor $\alpha$ şi $\beta$.

Aplicând relaţia raportului de corelaţie la datele din exemplul prezentat în acest capitol, se obţine:

$$\eta_{y_x} = \sqrt{\frac{0,5 \cdot 115 + 7,5 \cdot 420 - \frac{1}{5}(115)^2}{3225 - \frac{1}{5}(115)^2}} = \sqrt{\frac{562,5}{580}} = 0,98 .$$

Raportul de determinație ($\eta^2_{y_x} = 96\%$) estimează că variația variabilei *Y* este determinată în proporție de *96%* de factorul *X*, diferența până la *100%* s-ar datora factorilor aleatori.

Dacă valoarea la pătrat a raportului de corelație ($\eta^2_{y_x}$) este egală cu valoarea la pătrat a coeficientului de corelație empiric ($r^2$), conform **testului B (Blackman)**, $r^2 \approx \eta^2_{y_x}$, legătura este liniară.

### Testarea raportului de corelație

**Testul *F*** folosit pentru verificarea semnificației raportului de corelație este definit de relația:

$$F = \frac{n-k}{k-1} \cdot \frac{\hat{\eta}^2}{1-\hat{\eta}^2}$$

în care $n$ reprezintă numărul valorilor observate; $k$ – numărul parametrilor estimați ai modelului de regresie; $\hat{\eta}$ – estimatorul raportului de corelație.

Statistica $F$ urmează o lege de distribuție Snedecor-Fisher de $\nu_1 = k-1$ și $\nu_2 = n-k$ grade de libertate, unde $k$ este numărul parametrilor din model ($k = 2$ pentru modelul liniar simplu).

La nivelul unui eșantion observat, se determină **valoarea calculată a testului:** $F_{calc.} = \frac{n-k}{k-1} \cdot \frac{\eta^2_{y_x}}{1-\eta^2_{y_x}}$

Dacă $F_{calc.} > F_{tab.} = F_{\alpha,\nu_1,\nu_2}$ cu $\nu_1 = k-1$ și $\nu_2 = n-k$ grade de libertate, atunci se trage concluzia că variabila factorială influențează semnificativ comportarea variabilei rezultative, deci raportul de corelație diferă semnificativ statistic.

## 2.1.6. Regresia liniară simplă în SPSS

Procesul de estimare a parametrilor unui model de regresie în SPSS este cunoscut ca *„fitting the model"* sau *modelul estimat* și presupune

parcurgerea demersului: meniul **Analyze**→ comanda **Regression**→ opțiunea **Linear**, prin care se deschide fereastra de dialog **Linear Regression** (vezi Figura 5).

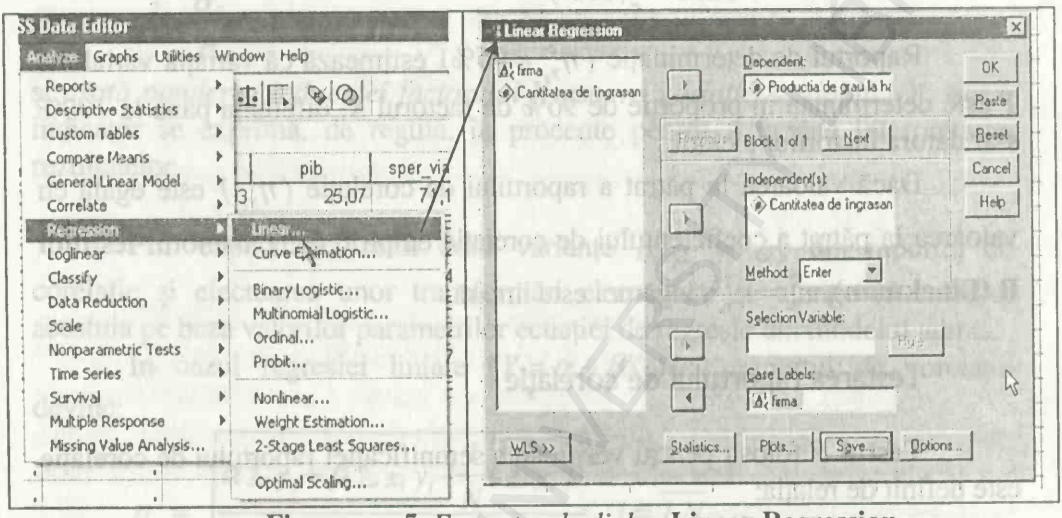



**Figura nr. 5.** *Fereastra de dialog* **Linear Regression**

În fereastra dialog **Linear Regression,** selectăm variabilele considerate şi le mutăm în zonele de lucru corespunzătoare. În exemplul nostru, selectăm variabila rezultativă ***prod*** şi o mutăm în zona **Dependent**, iar variabila factorială ***ingras*** în zona **Independent**. În zona **Case Labels** mutăm ***firma***.

**Tabelul nr. 4** *Cantitatea de îngrăşăminte şi producţia de grâu la ha*

| | firma | ingras | prod |
|---|---|---|---|
| 1 | a | 1,00 | 10,00 |
| 2 | b | 2,00 | 15,00 |
| 3 | c | 3,00 | 20,00 |
| 4 | d | 4,00 | 30,00 |
| 5 | e | 5,00 | 40,00 |

În continuare, se parcurg următorii paşi:

- Alegem din lista **Method**, ca metodă de lucru, opțiunea **Enter**;
- Activăm butonul de comandă **Statistics** care deschide fereastra de dialog **Linear Regression: Statistics** în care bifăm casetele de validare: **Estimates**, **Confidence intervals**, **Model fit** şi **Descriptives** (vezi Fig. 6);

- Butonul de comandă **Continue** determină revenirea în fereastra **Linear Regression** în care activăm butonul **Plots**, care deschide fereastra **Linear Regression: Plots**;
- În fereastra de dialog **Linear Regression: Plots** selectăm şi mutăm SRESID în zona **Y**, respectiv, ZPRED în zona **X**. Pentru **Standardized Residual Plots** bifăm casetele de validare **Histogram** şi **Normal probability plot**;
- Butonul de comandă **Continue** determină revenirea în fereastra **Linear Regression** în care activăm butonul **Save**;
- În fereastra **Linear Regression: Save** (vezi Fig.7), pentru **Predicted Values** bifăm caseta **Unstandardized**, pentru **Prediction Intervals** bifăm caseta **Mean**, iar pentru **Residuals** alegem **Unstandardized**;
- Acţionăm butonul de comandă **Continue** pentru a reveni în fereastra **Linear Regression**;
- Butonul **OK** comandă obţinerea output-ului în fereastra de rezultate şi a valorilor estimate în fişierul **Data Editor**.

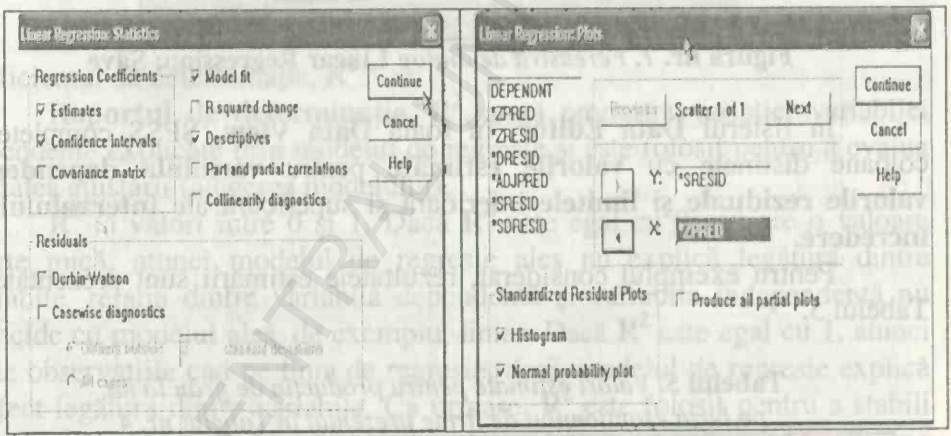



**Figura nr. 6.** *Ferestrele de dialog* **Statistics** *şi* **Plots** *pentru un model de regresie*

BCU IAŞI CENTRAL UNIVERSITY LIBRARY

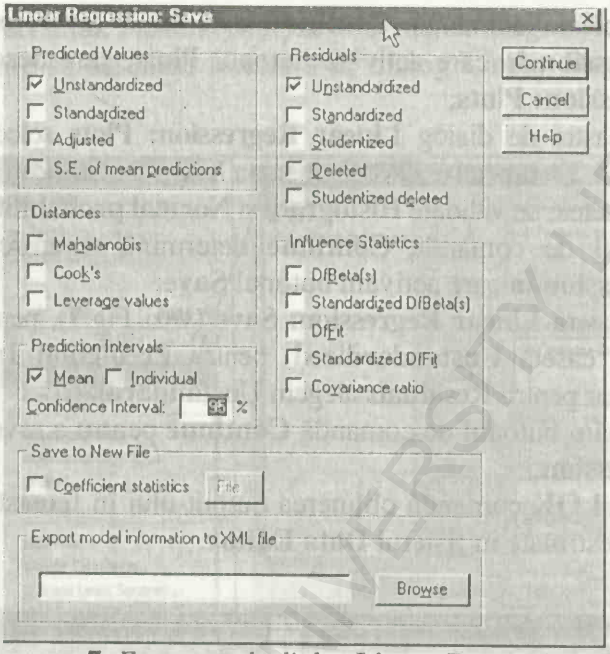



**Figura nr. 7.** *Fereastra de dialog* **Linear Regression: Save**

În fişierul **Data Editor**, în foaia **Data View,** SPSS completează coloane distincte cu **valorile estimate** pentru **variabila dependentă**, **valorile reziduale** şi **limitele** inferioară şi superioară ale **intervalului de încredere.**

Pentru exemplul considerat, rezultatele estimării sunt prezentate în Tabelul 5.

**Tabelul 5.** *Valori estimate pentru producţia de grâu la ha, pe baza eşantionului de firme prezentat în Tabelul nr. 4*

| firma | ingras | prod | pre_1 | res_1 | lmci_1 | umci_1 |
|---|---|---|---|---|---|---|
| a | 1.00 | 10.00 | 8.00000 | 2.00000 | 2.04619 | 13.95381 |
| b | 2.00 | 15.00 | 15.50000 | -.50000 | 11.29002 | 19.70998 |
| c | 3.00 | 20.00 | 23.00000 | -3.00000 | 19.56257 | 26.43743 |
| d | 4.00 | 30.00 | 30.50000 | -.50000 | 26.29002 | 34.70998 |
| e | 5.00 | 40.00 | 38.00000 | 2.00000 | 32.04619 | 43.95381 |

Fereastra de rezultate – **Output-ul,** pentru analiza de regresie, conţine: **Model Summary, ANOVA, Coefficients, Normal P-P plot** şi **Scatterplot.**

Tabelul **Model Summary** prezintă valoarea raportului de corelație (R), valoarea raportului de determinație ($R^2$), valoarea ajustată a lui R și eroarea standard a estimației. Pentru exemplul considerat, **Model Summary** este prezentat în Tabelul 6.

**Tabelul nr. 6.** *Model Summary, cazul regresiei simple*

| Model | R | R Square | Adjusted R Square | Std. Error of the Estimate |
|---|---|---|---|---|
| 1 | .985 | .970 | .960 | 2.4152 |

a Predictors: (Constant), Cantitatea de îngrășăminte la ha
b Dependent Variable: Producția de grâu la ha

**Valoarea R** arată dacă există sau nu o corelație între variabila dependentă (rezultativa Y) și variabila independentă (factoriala X). Acest indicator ia valori între 0 și 1.

**Interpretarea modelului.** În interpretarea modelului se folosește coeficientul de determinație, $R^2$.

**Raportul de determinație, $R^2$**, arată proporția variației variabilei dependente explicate prin modelul de regresie și este folosit pentru a evalua calitatea ajustării (alegerea modelului).

$R^2$ ia valori între 0 și 1. Dacă $R^2$ este egal cu 0 sau are o valoare foarte mică, atunci modelul de regresie ales nu explică legătura dintre variabile, relația dintre variabila dependentă și variabila independentă nu coincide cu modelul ales, de exemplu, liniar. Dacă $R^2$ este egal cu 1, atunci toate observațiile cad pe linia de regresie, deci, modelul de regresie explică perfect legătura dintre variabile. Ca urmare, $R^2$ este folosit pentru a stabili care model de regresie este cel mai bun. Această metodă de alegere a modelului de regresie potrivit este recomandată pentru modelele care nu conțin un număr mare de variabile.

Pentru exemplul considerat a rezultat o valoare **R = 0.985**, respectiv, **$R^2$= 0.970**, ceea ce ne arată că între cantitatea de producție/ha și cantitatea de îngrășăminte/ha există o legătură liniară, directă, foarte strânsă.

**Tabelul Regression ANOVA** prezintă rezultatele analizei variației variabilei dependente sub influența factorului de regresie și a factorului reziduu. Adică, prezintă informații asupra sumei pătratelor abaterilor variabilei dependente, datorate modelului de regresie și factorului reziduu,

gradele de libertate, estimațiile varianțelor datorate celor două surse de variație (regresie și reziduu), raportul F și Sig. (vezi Tabelul 7).

**Tabelul nr. 7.** *ANOVA pentru regresie*

| Model | | Sum of Squares | df | Mean Square | F | Sig. |
|---|---|---|---|---|---|---|
| 1 | Regression | 562.500 | 1 | 562.500 | 96.429 | .002 |
| | Residual | 17.500 | 3 | 5.833 | | |
| | Total | 580.000 | 4 | | | |

a Predictors: (Constant), Cantitatea de îngrășăminte la ha
b Dependent Variable: Producția de grâu la ha

**Statistica test F** se obține ca raport între media pătratelor abaterilor datorate regresiei și media pătratelor abaterilor datorate reziduului, calculate cu gradele de libertate corespunzătoare. Această statistică test este folosită pentru testarea modelului de regresie, adică a ipotezei prin care se presupune că panta dreptei ($\beta_1$) este 0, respectiv, pentru regresia multiplă, $\beta_1, \ldots \beta_p = 0$.

Dacă **testul F** ia o valoare mare, iar valoarea **Sig.** corespunzătoare statisticii **F** este mică (mai mică decât 0,05), atunci variabila independentă explică variația variabilei dependente și invers.

În exemplul considerat, valoarea **Sig.** pentru **F** este mai mică decât 0,05, deci relația liniară dintre cele două variabile considerate este semnificativă (vezi Tabelul 7).

**Tabelul nr. 8.** *Coeficienții de regresie*

| | | Unstandardized Coefficients | | Standardized Coefficients | t | Sig. |
|---|---|---|---|---|---|---|
| Model | | B | Std. Error | Beta | | |
| 1 | (Constant) | .500 | .533 | | .197 | .856 |
| | Cantitatea de îngrășăminte la ha | 7.500 | .764 | .985 | .820 | .002 |

a Dependent Variable: Producția de grâu la ha

## Coeficienții de regresie

Tabelul **Coefficients** (vezi Tabelul 8) prezintă coeficienții nestandardizați ai modelului de regresie estimat, erorile standard ale acestora, coeficienții de regresie standardizați cu erorile standard

corespunzătoare, precum şi valorile statisticii test **t** şi valorile **Sig.** corespunzătoare.

Coeficienţii de regresie standardizaţi sunt folosiţi atunci când într-un model intră mai multe variabile independente exprimate în unităţi de măsură diferite, în scopul facilitării comparării acestora.

Testarea parametrilor modelului de regresie se face cu ajutorul testului t, pentru a afla dacă aceştia diferă semnificativ de zero:

$$H_0 : \beta = 0$$

Pentru exemplul dat, valoarea (*Sig.=0.002*) este mai mică decât 0.05, arătând că $\beta$ (panta dreptei de regresie) este semnificativ diferit de zero şi corespunde unei legături semnificative între cele două variabile.

## 2.2. Modelul neliniar simplu

### 2.2.1. Formularea problemei şi exemple din economie

O mare parte a legăturilor dintre variabilele din domeniul economic nu sunt de tip liniar, ci se încadrează în următoarele modele: parabolic, hiperbolic, exponenţial, polinomial etc. O parte dintre aceste modele se aduc la forma modelului liniar prin logaritmare şi se tratează ca atare.

Un exemplu clasic de model neliniar este modelul care explică relaţia dintre inflaţie şi şomaj, iar curba reprezentată cu ajutorul celor două variabile se numeşte *curba lui Philips*. Curba are la bază un model reciproc sau hiperbolic de forma:

$$r_i = \alpha + \beta \frac{1}{s_i} + \varepsilon_i ,$$

unde, $r$ este rata inflaţiei, iar $s$ este rata şomajului.

## 2.2.2. Modele liniarizabile

Modelele liniarizabile sunt modelele neliniare care permit, în urma unor transformări, construirea de modele liniare. Astfel de modele sunt: exponențial, hiperbolic, putere ş.a.

**Ecuația estimată a modelului de regresie de tip hiperbolic** este:

$$Y_x = a + \frac{b}{X}$$

Estimațiile parametrilor modelului se află pe baza sistemului:

$$na + b\sum \frac{1}{x_i} = \sum y_i$$

$$a\sum \frac{1}{x_i} + b\sum \frac{1}{x_i^2} = \sum \frac{y_i}{x_i}$$

**Ecuația estimată a modelului de regresie de tip exponențial** are forma:

$$Y_x = a \cdot b^X$$

Pentru orice eşantion, expresia poate fi liniarizată prin logaritmare:

$$\log Y_x = \log a + X \cdot \log b$$

Aplicând metoda celor mai mici pătrate, pentru un eşantion observat se obține sistemul:

$$n \cdot \log a + \log b \cdot \sum x_i = \sum \log y_i$$

$$\log a \cdot \sum x_i + \log b \cdot \sum x_i^2 = \sum x_i \cdot \log y_i$$

Rezolvarea sistemului de mai sus permite aflarea relațiilor prin care se determină estimațiile modelului.

*Observație*

În practică, utilizând pachetele program de statistică, pentru estimarea parametrilor unui model de tip exponențial se utilizează un model de forma:

$$\log Y = \alpha + \beta X + \varepsilon$$

BCU IAŞI CENTRU UNIVERSITY LIBRARY

În aceste condiții, panta dreptei semnifică variația relativă a variabilei dependente, la o variație absolută de o unitate a variabilei independente: $\beta = \frac{d \ln Y}{dX}$.

**Ecuația estimată a modelului de regresie de tip putere** are forma:

$Y_x = a \cdot X^b$

Pentru orice eşantion, expresia poate fi liniarizată prin logaritmare:

$\log Y_x = \log a + b \cdot \log X \cdot$

Aplicând metoda celor mai mici pătrate, pentru un eşantion observat se obține sistemul:

$$n \cdot \log a + b \cdot \sum \log x_i = \sum \log y_i$$

$$\log a \cdot \sum \log x_i + b \cdot \sum (\log x_i)^2 = \sum \log x_i \cdot \log y$$

*Observație*

În practică, utilizând pachetele program de statistică, pentru estimarea parametrilor unui model de tip putere se utilizează un model de forma:

$$\log Y = \alpha + \beta \log X + \varepsilon$$

## Elasticitatea

Elasticitatea unei variabile $Y$ în raport cu o altă variabilă $X$ reprezintă modificarea relativă (procentuală) a variabile $Y$ la o modificare relativă (procentuală) dată a lui $X$, de obicei mică, de o unitate.

Formalizând, elasticitatea $E$ este dată prin relația:

$$E = \frac{\%\, modif.\, Y}{\%\, modif.\, X} = \frac{\frac{\Delta Y}{Y} 100}{\frac{\Delta X}{X} 100} = \frac{\Delta Y}{\Delta X} \frac{X}{Y},$$

unde operatorul $\Delta$ semnifică modificările sau diferențele realizate la nivelul unei variabile (operatorul diferențial).

Pentru variații fine ale variabilelor, elasticitatea se poate aproxima prin relația:

$$E = \frac{dY}{dX} \frac{X}{Y} = \frac{d \ln Y}{d \ln X}$$

*Observaţie*

Pentru un model de regresie de tip putere, considerat în forma liniarizată, $\log Y = \alpha + \beta \log X + \varepsilon$, parametrul $\beta$ este tocmai elasticitatea variabilei dependente $Y$ în raport cu variabila independentă $X$:

$$\beta = \frac{d \ln Y}{d \ln X} = E.$$

### 2.2.3. Modele polinomiale

Cel mai simplu model polinomial este modelul parabolic.

**Ecuaţia estimată a modelului de regresie de tip parabolic**

$$Y_x = a + bX + cX^2$$

Pentru calculul estimaţiilor parametrilor la nivelul unui eşantion, aplicând metoda celor mai mici pătrate, se rezolvă sistemul:

$$na + b\sum x_i + c\sum x_i^2 = \sum y_i$$

$$a\sum x_i + b\sum x_i^2 + c\sum x_i^3 = \sum x_i y_i$$

$$a\sum x_i^2 + b\sum x_i^3 + c\sum x_i^4 = \sum x_i^2 y_i$$

**Exemplu.** Admitem că între numărul de pomi la hectar ($X$) şi producţia medie la hectar ($Y$) există o legătură. Pentru a cunoaşte legătura dintre cele două variabile s-au înregistrat datele cuprinse în tabelul 9.

**Se cere:**

**1.** Să se stabilească forma şi direcţia legăturii dintre cele două variabile prin metoda grafică;

**2.** Să se determine valorile ajustate ale lui $Y$ după funcţia de regresie corespunzătoare legăturii;

**3.** Să se calculeze intensitatea legăturii dintre variabilele admise.

**Rezolvare**

**1.** Corelograma, reprezentând legătura dintre numărul de pomi la hectar ($X$) şi producţia medie la hectar ($Y$), este prezentată în figura 8.

Reprezentarea grafică a celor două variabile arată o legătură directă de forma unei parabole.

BCU ALECU RUSSO UNIVERSITY LIBRARY

**2.** Ecuația de regresie corespunzătoare legăturii dintre cele două variabile este de forma:

$$Y_x = a + bX + cX^2$$

**Tabelul nr. 9.** *Date convenționale*

| Nr. crt. | X | Y |
|---|---|---|
| 1 | 2 | 3 |
| 1 | 20 | 21,2 |
| 2 | 25 | 23,0 |
| 3 | 30 | 24,3 |
| 4 | 35 | 28,8 |
| 5 | 40 | 27,5 |
| 6 | 45 | 28,7 |
| 7 | 50 | 29,0 |
| 8 | 55 | 28,5 |
| 9 | 60 | 27,8 |

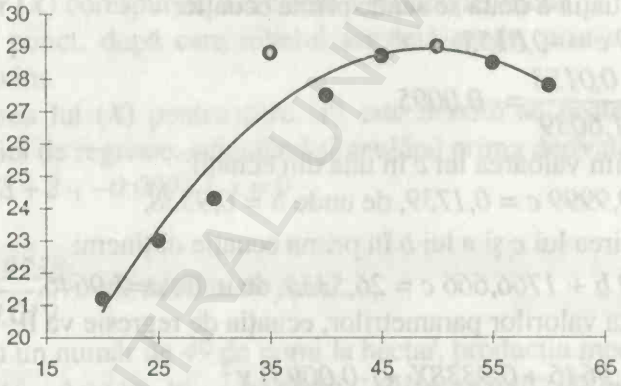



**Figura nr. 8.** *Legătura dintre numărul de pomi la hectar și producția medie la hectar*

Pentru determinarea valorilor ajustate ale lui *Y*, se estimează parametrii ecuației prin rezolvarea sistemului:

$$na + b\sum x_i + c\sum x_i^2 = \sum y_i$$

$$a\sum x_i + b\sum x_i^2 + c\sum x_i^3 = \sum x_i y_i$$

$$a\sum x_i^2 + b\sum x_i^3 + c\sum x_i^4 = \sum x_i^2 y_i$$

Sumele din sistem se obțin cu ajutorul tabelului de calcul nr. 10. Înlocuind sumele se obține sistemul:

$$9a + 360b + 15900c = 238,8$$

BCU "LUCIAN BLAGA" CENTRAL UNIVERSITY LIBRARY

$360\,a + 15900\,b + 756000\,c = 9813$
$15900\,a + 756000\,b + 37882500\,c = 440915$

Pentru rezolvarea acestui sistem, se poate folosi procedeul excluderii succesive a primului coeficient. În acest scop, se împarte prima ecuație cu *9*, a doua ecuație cu *360* și a treia ecuație cu *15900* și se obține:

$a + 40\,b + 1766{,}6667\,c = 26{,}5333$
$a + 44{,}1667\,b + 2100\,c = 27{,}2583$
$a + 47{,}5472\,b + 2382{,}5471\,c = 27{,}7305$

Din ecuația a doua și ecuația a treia se scade prima ecuație:

$4{,}1667\,b + 333{,}3334\,c = 0{,}7250$
$7{,}5472\,b + 615{,}8804\,c = 1{,}1972$

Se împarte prima ecuație cu *4,1667*, iar a doua cu *7,5472* și se obține:

$b + 79{,}9999\,c = 0{,}1739$
$b + 81{,}6938\,c = 0{,}1586$

Din ecuația a doua se scade prima ecuație:

$1{,}6039\,c = -0{,}0153$

de unde, $c = \dfrac{-0{,}0153}{1{,}6039} = -0{,}0095$

Înlocuim valoarea lui *c* în una din ecuații:

$b + 79{,}9999\,c = 0{,}1739$, de unde $b = 0{,}9338$,

iar prin înlocuirea lui *c* și a lui *b* în prima ecuație obținem:

$a + 40\,b + 1766{,}666\,c = 26{,}5333$, de unde $a = 5{,}9646$.

Pe baza valorilor parametrilor, ecuația de regresie va fi:

$$Y_x = 5{,}9646 + 0{,}9338X + (-0{,}0095)X^2$$

**Tabelul nr. 10.** *Elemente de calcul*

| $X$ | $Y$ | $x^2$ | $x^3$ | $x^4$ |
|---|---|---|---|---|
| 1 | 2 | 3 | 4 | 5 |
| 20 | 21,2 | 400 | 8000 | 160000 |
| 25 | 23,0 | 625 | 15625 | 390625 |
| 30 | 24,3 | 900 | 27000 | 810000 |
| 35 | 28,8 | 1225 | 42875 | 1500625 |
| 40 | 27,5 | 1600 | 64000 | 2560000 |
| 45 | 28,7 | 2025 | 91125 | 4100625 |
| 50 | 29,0 | 2500 | 125000 | 6250000 |
| 55 | 28,5 | 3025 | 166375 | 9150625 |
| 60 | 27,8 | 3600 | 216000 | 12960000 |
| 360 | 238,8 | 15900 | 756000 | 37882500 |

BCU IAȘI - BIBLIOTECA UNIVERSITY LIBRARY

| $X$ | $xy$ | $x^2 y$ | $y_x = a + bx + cx^2$ | $y^2$ |
|---|---|---|---|---|
| 1 | 6 | 7 | 8 | 9 |
| 20 | 424,0 | 8480,0 | 21,0426 | 449,44 |
| 25 | 575,0 | 14375 | 23,3721 | 529,0 |
| 30 | 729,0 | 21870 | 25,4286 | 520,49 |
| 35 | 1008,0 | 35280 | 27,0101 | 829,44 |
| 40 | 1100,0 | 44000 | 28,1166 | 756,25 |
| 45 | 1291,5 | 58117,5 | 28,7481 | 823,69 |
| 50 | 1450,0 | 72500 | 28,9046 | 841,00 |
| 55 | 1567,5 | 86212,5 | 28,5861 | 812,25 |
| 60 | 1668,0 | 100080,0 | 27,7926 | 772,84 |
| 360 | 9813 | 440915 | 239 | 6404,4 |

Semnul parametrilor de regresie arată că la o creştere a numărului de pomi la hectar ($X$) corespunde o creştere a producţiei medii la hectar ($Y$) până la un anumit punct, după care nivelul acesteia scade, prin urmare, ecuaţia admite un maxim.

Valoarea lui ($X$) pentru care ($Y$) este maxim se poate determina cu ajutorul ecuaţiei de regresie, calculând şi anulând prima derivată:

$$0,9338 + 2 \cdot (-0.0095) \cdot x = 0$$

de unde,

$$x = \frac{0,9338}{0,019} = 49,1473 \approx 49 \text{ pomi}.$$

Pentru un număr de *49* de pomi la hectar, producţia medie la hectar va fi: $y_x = 5,9646 + 0,9338 \cdot 49 + (-0,0095) \cdot 49^2 = 28,9303$ unităţi de producţie la hectar.

Pentru un număr mai mare de *49* pomi la hectar, producţia medie scade. Astfel, pentru *50* pomi producţia este:

$y_x = 5,9646 + 0,9338 \cdot 50 + (0,0095) \cdot 50^2 = 28,9046$ unităţi de producţie la hectar.

3. Intensitatea legăturii dintre cele două variabile se stabileşte cu ajutorul raportului de corelaţie, calculat după relaţia:

BCU "ASE" CENTRAL UNIVERSITY LIBRARY

$$\eta_{y_x} = \sqrt{\frac{a\sum y_i + b\sum x_i y_i + c\sum x_i^2 y_i - \frac{1}{n}(\sum y_i)^2}{\sum y_i^2 - \frac{1}{n}(\sum y_i)^2}} =$$

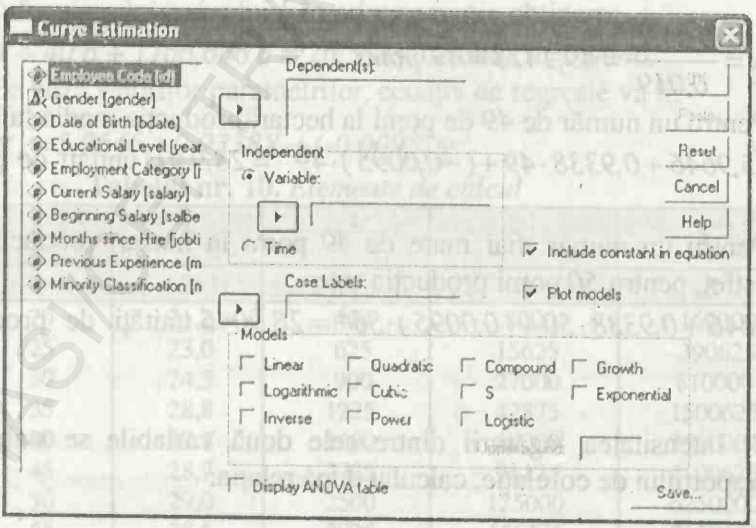
$$= \sqrt{\frac{5{,}9646 \cdot 238{,}8 + 0{,}9338 \cdot 9813 + (-0{,}0095) \cdot 440915 - \frac{1}{9}(238{,}8)^2}{6404{,}4 - \frac{1}{9}(238{,}8)^2}} =$$

$$= \sqrt{0{,}8139} = 0{,}9021$$

Valoarea raportului de corelație arată că între variabilele admise în studiu există o corelație foarte strânsă, iar valoarea raportului de determinație arată că *81,39%* din variația producției medii s-ar datora densității pomilor la hectar.

## Modelele de regresie simplă în SPSS

În pachetul SPSS, există posibilitatea construirii unui număr de modele de regresie simplă, utilizând meniul ***Analyze – Regression – Curve Estimation.***



**Modelele disponibile în SPSS sunt:**

**1. Modelul liniar simplu**

$$Y = \alpha + \beta X + \varepsilon$$

**2. Modelul logaritmic**

$$Y = \alpha + \beta \ln X + \varepsilon$$

**3. Modelul reciproc, invers**

$$Y = \alpha + \beta \frac{1}{X} + \varepsilon$$

**4. Modelul quadratic**

$$Y = \alpha + \beta_1 X + \beta_2 X^2 + \varepsilon$$

**5. Modelul cubic**

$$Y = \alpha + \beta_1 X + \beta_2 X^2 + \beta_3 X^3 + \varepsilon$$

**6. Modelul Power**

$$Y_X = \alpha \cdot X^{\beta}$$

În formă liniarizată, se construieşte **modelul de regresie**:

$$\ln Y = \ln \alpha + \beta \cdot \ln X + \varepsilon$$

**7. Modelul Compound**

$$Y_X = \alpha \cdot \beta^X$$

În formă liniarizată, se construieşte **modelul de regresie**:

$$\ln Y = \ln \alpha + \ln \beta \cdot X + \varepsilon$$

**8. Modelul *S***

$$Y_X = e^{\alpha + \beta \frac{1}{X}}$$

În formă liniarizată, se construieşte **modelul de regresie**:

$$\ln Y = \alpha + \beta \cdot \frac{1}{X} + \varepsilon$$

BCU IASI / CENTRAL UNIVERSITY LIBRARY

**9. Modelul Logistic**

$$Y_X = \frac{1}{\frac{1}{k} + \alpha \cdot \beta^X}$$

În formă liniarizată, se construieşte **modelul de regresie:**

$$ln(\frac{1}{Y} - \frac{1}{k}) = ln\,\alpha + ln\,\beta \cdot X + \varepsilon$$

**10. Modelul Growth**

$$Y_X = e^{\alpha + \beta X}$$

În formă liniarizată, se construieşte **modelul de regresie:**

$$ln\,Y = \alpha + \beta \cdot X + \varepsilon$$

**11. Modelul exponenţial**

$$Y_X = \alpha \cdot e^{\beta X}$$

În formă liniarizată, se construieşte **modelul de regresie:**

$$ln\,Y = ln\,\alpha + \beta \cdot X + \varepsilon$$

BCU IASI/CENTRAL UNIVERSITY LIBRARY

Econometrie

# Capitolul 3

# MODELE DE REGRESIE MULTIPLĂ

1. Modelul liniar multiplu
   1.1. Prezentare model şi ipoteze
   1.2. Estimarea parametrilor modelului
   1.3. Testarea parametrilor modelului
   1.4. Coeficienţi de corelaţie
   1.5. Regresia multiplă în SPSS
2. Modele neliniare multiple
   2.1. Prezentarea problemei. Exemple din economie
   2.2. Modele liniarizabile
   2.3. Modele polinomiale

În modelarea econometrică, cele mai frecvent utilizate sunt modelele de regresie multiplă, care conţin cel puţin două variabile independente.

## 3.1. Modelul liniar multiplu

Modelul liniar multiplu este un model în care între variabila dependentă şi variabilele independente există o legătură liniară.

BCU IASI / CENTRAL UNIVERSITY LIBRARY

### 3.1.1. Prezentare model şi ipoteze

Un model statistic de regresie liniară multiplă este definit de relaţia:

$$Y = \alpha + \beta_1 X_1 + \beta_2 X_2 + ... + \beta_p X_p + \varepsilon$$

unde:
- $Y$ este variabila dependentă ;
- $X_1, X_2, ..., X_p$ sunt variabile independente (predictori);
- $k$ este numărul de parametri din model, $k=p+1$;
- $\varepsilon$ este variabila aleatoare eroare (reziduu);
- $\alpha, \beta_i$ sunt coeficienţii de regresie.

Modelarea se realizează în următoarele ipoteze:
- normalitatea erorilor la nivelul distribuţiilor condiţionate de tipul $Y_i|X = x_i$ : $\varepsilon_i \sim N(0, \sigma^2)$, adică variabila reziduală urmează o lege de repartiţie normală de medie zero şi varianţă $\sigma^2$;
- homoscedasticitate: $V(\varepsilon_i) = M(\varepsilon_i^2) = \sigma^2$, adică varianţa erorii este constantă la nivelul distribuţiilor condiţionate de tipul $Y_i|X = x_i$ ;
- necorelarea erorilor: $cov(\varepsilon_i, \varepsilon_j) = 0$, adică erorile nu se influenţează reciproc;
- lipsa corelaţiei dintre variabilele independente şi variabila eroare;
- lipsa coliniarităţii sau a unei legături liniare între variabilele independente.

### 3.1.2. Estimarea parametrilor modelului

Ecuaţia estimată a modelului de regresie care exprimă o legătură multiplă liniară este:

$$Y_{x_1 x_2 ...} = a + b_1 X_1 + b_2 X_2 + ... b_p X_p$$

Metoda celor mai mici pătrate presupune condiţia de minim:

$$S = \sum (y_i - y_{x_1 x_2 ...})^2 = \sum (y_i - a - b_1 x_{1i} - b_2 x_{2i} - ... - b_p x_{pi})^2 = minim$$

Dacă anulăm derivatele parţiale ale expresiei de mai sus în raport cu parametrii $a, b_1, b_2, ..., b_p$, rezultă sistemul de ecuaţii normale prin a cărui rezolvare se obţin estimaţiile parametrilor ecuaţiei de regresie.

De exemplu, pentru un model cu două variabile independente ($p=2$), $Y=\alpha+\beta_1 X_1+\beta_2 X_2+\varepsilon$, sistemul de ecuații normale, la nivelul unui eşantion, este de forma:

$$\begin{cases} na+b_1\sum_i x_{1i}+b_2\sum_i x_{2i}=\sum_i y_i \\ a\sum_i x_{1i}+b_1\sum_i x_{1i}^2+b_2\sum_i x_{1i}x_{2i}=\sum_i y_i x_{1i} \\ a\sum_i x_{2i}+b_1\sum_i x_{1i}x_{2i}+b_2\sum_i x_{2i}^2=\sum_i y_i x_{2i} \end{cases}$$

Prin rezolvarea sistemului, se obțin relații pentru estimațiile parametrilor modelului de regresie.

### 3.1.3. Testarea parametrilor modelului

Considerăm modelul de regresie liniară multiplă:

$Y=\alpha+\beta_1 X_1+\beta_2 X_2+\varepsilon$

Prin aplicarea metodei celor mai mici pătrate, pentru coeficienții de regresie se obțin estimatori care urmează legea normală:

$\hat{\beta}_i \sim N(\beta_i, \sigma^2_{\hat{\beta}_i})$, $i=\overline{1,p}$.

Prin standardizare şi aplicarea teoremei limită centrală, se obțin statisticile:

$\frac{\hat{\beta}_i-\beta_i}{\sigma_{\hat{\beta}_i}} \sim N(0,1)$, respectiv, $\frac{\hat{\beta}_i-\beta_i}{\hat{\sigma}_{\hat{\beta}_i}} \sim t(n-k)$,

unde $k$ este numărul de parametri din model. Pentru un model cu două variabile independente, $k=3$.

Testarea semnificației parametrilor $\beta_i$ se realizează cu ajutorul testului student:

$$t=\frac{\hat{\beta}_i-\beta_i}{\hat{\sigma}_{\hat{\beta}_i}}.$$

BCU IAŞI CENTRAL UNIVERSITY LIBRARY

### Etapele testării

Considerăm un model cu două variabile independente. Testarea presupune parcurgerea demersului:

- Ipoteze: $H_0 : \beta_i = 0$, $i = \overline{1,2}$, respectiv $\alpha = 0$

$H_1 : \beta_i \neq 0$, respectiv $\alpha \neq 0$

- Pragul de semnificație, de regulă, $\alpha = 0{,}05$;
- Statistica Student, în condițiile acceptării ipotezei nule:

$t = \frac{\hat{\beta}_i}{\hat{\sigma}_{\hat{\beta}_i}} \sim t(n-3)$, respectiv $t = \frac{\hat{\alpha}}{\hat{\sigma}_{\hat{\alpha}}} \sim t(n-3)$.

De exemplu, pentru $\hat{\beta}_1$, estimatorul variației este:

$\hat{\sigma}^2_{\hat{\beta}_1} = \frac{\hat{\sigma}^2_{\hat{\varepsilon}}}{\sum_i (x_{1i} - \bar{x}_1)^2 (1 - r^2_{x_1 x_2})}$, unde $r^2_{x_1 x_2}$ este coeficientul de determinație dintre variabilele independente, iar

$\hat{\sigma}^2_{\hat{\varepsilon}} = \frac{\sum_i \left(y_i - \hat{\alpha} - \hat{\beta}_1 x_{1i} - \hat{\beta}_2 x_{2i}\right)^2}{n-3}$;

- valorile teoretice $t_{(n-3),\alpha/2}$, care se citesc din tabela Student;
- valoarea calculată la nivelul unui eșantion:

pentru coeficienții de regresie - $t_{calc.} = \frac{b_i}{s_{\hat{\beta}_i}}$,

pentru ordonata la origine: $t_{calc.} = \frac{a}{s_{\hat{\alpha}}}$;

- regula de decizie: $t_{calc.} \in [-t_{(n-3),\alpha/2}; t_{(n-3),\alpha/2}]$ se acceptă $H_0$, în rest se respinge ipoteza nulă, cu probabilitatea $(1 - \alpha)$.

### Abordare matriceală

Modelul de regresie liniară multiplă se poate scrie sub formă matriceală prin relația: $Y = X \cdot \beta + \varepsilon$, unde

$$Y = \begin{pmatrix} Y_1 \\ Y_2 \\ \dots \\ Y_n \end{pmatrix}, \qquad X = (1\ X_1\ X_2 \dots X_p) = \begin{pmatrix} 1\ x_{11}\ x_{21} \dots x_{p1} \\ 1\ x_{12}\ x_{22} \dots x_{p2} \\ \dots\ \dots\ \dots\ \dots \\ 1\ x_{1n}\ x_{2n} \dots x_{pn} \end{pmatrix},$$

$$\beta = \begin{pmatrix} \alpha \\ \beta_1 \\ \dots \\ \beta_p \end{pmatrix}, \qquad \varepsilon = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \dots \\ \varepsilon_n \end{pmatrix}.$$

*Observaţie*

Prima coloană din matricea corespunzătoare valorilor variabilelor independente este coloana corespunzătoare parametrului $\alpha$, ale cărei valori sunt egale cu unu.

Pentru două variabile independente, $p=2$ sau $k = 3$, avem modelul de regresie multiplă de forma:

$$y_i = M(Y / X\ ) + \varepsilon_i = \alpha + \beta_1 x_{1i} + \beta_2 x_{2i} + \varepsilon_i$$

Scris matriceal, modelul de mai sus este dat prin: $Y = X \cdot \beta + \varepsilon$, unde

$$Y = \begin{pmatrix} Y_1 \\ Y_2 \\ \dots \\ Y_n \end{pmatrix}, \qquad X = (1\ X_1\ X_2) = \begin{pmatrix} 1\ x_{11}\ x_{21} \\ 1\ x_{12}\ x_{22} \\ \dots\ \dots \\ 1\ x_{1n}\ x_{2n} \end{pmatrix},$$

$$\beta = \begin{pmatrix} \alpha \\ \beta_1 \\ \beta_2 \end{pmatrix}, \qquad \varepsilon = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \dots \\ \varepsilon_n \end{pmatrix}.$$

## Rezolvarea ecuaţiei matriceale

Scris matriceal, sistemul de ecuaţii rezultat din metoda celor mai mici pătrate are forma: $X' \cdot X \cdot \hat{\beta} = X' \cdot Y$, adică sub forma unei ecuaţii matriceale a cărei soluţie este:

BCU "SPIRU HARET" UNIVERSITY LIBRARY

$$\hat{\beta} = (X' \cdot X)^{-1} X' \cdot Y$$

### Proprietăţile estimatorilor

Proprietăţile estimatorilor parametrilor modelului de regresie se pot determina pe baza relaţiei:

$$\hat{\beta} = (X' \cdot X)^{-1} X' \cdot Y = (X' \cdot X)^{-1} X' (X \cdot \beta + \varepsilon) = \beta + (X' \cdot X)^{-1} X' \cdot \varepsilon .$$

Rezultă proprietăţile:

- nedeplasare, $M(\hat{\beta}) = \beta$
- matricea varianţei estimatorilor este $\Omega_{\hat{\beta}} = \sigma^2 (X' \cdot X)^{-1}$,

  unde $\varepsilon_i \sim N(0, \sigma^2)$, iar un estimator al dispersiei erorilor este dat prin

  $$\hat{\sigma}_{\hat{\varepsilon}}^2 = \frac{\sum_i \hat{\varepsilon}_i^2}{n-3} = \frac{\sum_i (Y_i - \hat{\alpha} - \hat{\beta}_1 X_{1i} - \hat{\beta}_2 X_{2i})^2}{n-3};$$

- estimatori normal distribuiţi, $\hat{\beta}_i \sim N(\beta_i, \sigma_{\hat{\beta}_i}^2)$. Prin standardizare se obţin statisticile: $\frac{\hat{\beta}_i - \beta_i}{\sigma_{\hat{\beta}_i}} \sim N(0,1)$, iar dacă parametrii $\sigma_{\hat{\beta}_i}^2$ nu sunt cunoscuţi, se construiesc variabilele Student:

  $$\frac{\hat{\beta}_i - \beta_i}{\hat{\sigma}_{\hat{\beta}_i}} \sim t(n-3);$$

- intervalele de încredere pentru parametrii modelului de regresie sunt: $(\hat{\beta}_i \pm t_{\alpha/2} \hat{\sigma}_{\hat{\beta}_i})$,

  iar la nivelul unui eşantion: $(b_i \pm t_{\alpha/2} s_{\hat{\beta}_i})$

## 3.1.4. Coeficienţi de corelaţie

Măsurarea intensităţii corelaţiei multiple se poate efectua cu ajutorul următorilor indicatori: coeficientului de corelaţie multiplă (doar pentru modelul liniar), raportul de corelaţie multiplă, raportul de determinaţie multiplă, coeficienţi de corelaţie parţială şi coeficienţi de determinaţie parţială.

BCU "ION CREANGĂ" UNIVERSITY LIBRARY

## Coeficientul de corelaţie multiplă

Coeficientul de corelaţie multiplă se determină cu ajutorul coeficienţilor de corelaţie simplă dintre variabilele perechi. Astfel, în cazul corelaţiei dintre o variabilă rezultativă $Y$ şi două variabile independente $X_1$, $X_2$, la nivelul unui eşantion, *coeficientul de corelaţie multiplă*, notat cu $R$, se calculează după relaţia:

$$R = \sqrt{\frac{r_{yx_1}^2 + r_{yx_2}^2 - 2\, r_{yx_1} r_{yx_2} r_{x_1x_2}}{1 - r_{x_1x_2}^2}},$$

în care:

$$r_{yx_1} = \frac{n\sum x_1 y - \sum x_1 \sum y}{\sqrt{[n\sum x_1^2 - (\sum x_1)^2][n\sum y^2 - (\sum y)^2]}}$$

$$r_{yx_2} = \frac{n\sum x_2 y - \sum x_2 \sum y}{\sqrt{[n\sum x_2^2 - (\sum x_2)^2][n\sum y^2 - (\sum y)^2]}}$$

$$r_{x_1x_2} = \frac{n\sum x_1 x_2 - \sum x_1 \sum x_2}{\sqrt{[n\sum x_1^2 - (\sum x_1)^2][n\sum x_2^2 - (\sum x_2)^2]}}$$

Dacă $r_{x_1x_2} = 0$, , atunci $R = \sqrt{r_{yx_1}^2 + r_{yx_2}^2}$.

Coeficientul de corelaţie multiplă liniară *se poate determina şi cu ajutorul unei relaţii bazate pe determinanţii formaţi din coeficienţii de corelaţie simplă liniară*. Astfel, în cazul a trei variabile, vom avea:

$$R = \sqrt{1 - \frac{\Delta y_{y_x}}{\Delta y}}$$

unde:

$$\Delta_y = \begin{vmatrix} 1 & r_{x_1x_2} \\ r_{x_1x_2} & 1 \end{vmatrix}; \quad \Delta y_{y_x} = \begin{vmatrix} 1 & r_{yx_1} & r_{yx_2} \\ r_{x_1y} & 1 & r_{x_1x_2} \\ r_{x_2y} & r_{x_2x_1} & 1 \end{vmatrix}$$

## Raportul de corelaţie multiplă

La nivelul unui eşantion, raportul de corelaţie multiplă se calculează după relaţia:

$$\eta_{yx_1x_2\ldots} = \sqrt{\frac{s^2_{yx_1x_2\ldots}}{s^2_y}}$$

sau $$\eta_{yx_1x_2\ldots} = \sqrt{1 - \frac{s^2_{y/yx_1x_2\ldots}}{s^2_y}} = \sqrt{1 - \frac{\sum(y_i - y_{x_1x_2\ldots})^2}{\sum(y_i - \bar{y})^2}}$$

Raportul de corelație multiplă se poate calcula și pe baza estimațiilor parametrilor ecuației de regresie multiplă, formula stabilindu-se în funcție de modelul regresiei multiple aplicat.

Pentru o corelație multiplă liniară dintre $Y$ și $X_1, X_2$, a cărui model estimat este de forma:

$$y_{x_1x_2} = a + b_1x_1 + b_2x_2,$$

raportul de corelație devine:

$$\eta_{yx_1x_2} = \sqrt{\frac{a\sum y + b_1\sum x_1y + b_2\sum x_2y - \frac{1}{n}(\sum y)^2}{\sum y^2 - \frac{1}{n}(\sum y)^2}}$$

## Corelația parțială

Alături de corelația simplă, care arată legătura dintre o variabilă rezultativă ($Y$) și o variabilă factorială ($X$) și de corelația multiplă, care caracterizează legătura dintre $Y$ și variația simultană a două sau mai multe variabile factoriale, în practică apare necesitatea studierii separate a perechilor de variabile $Y$ și $X$. Aceasta se realizează cu ajutorul *corelației parțiale, care măsoară dependența dintre variabile prin excluderea succesivă a influenței celorlalți factori (considerând influența lor constantă), menținând numai influența factorului măsurat.*

În funcție de numărul variabilelor a căror influență se elimină din calcul, *coeficienții de corelație parțială pot fi de ordinul întâi (pentru o variabilă eliminată), de ordinul doi (pentru două variabile)* etc.

Coeficienții corelației parțiale pot fi calculați fie pe baza coeficienților de corelație simpli, fie pe baza dispersiilor, așa cum rezultă din relațiile de mai jos.

*Coeficienții de corelație parțială de ordinul întâi:*

- între $Y$ și $X_1$, excluzând influența lui $X_2$: $r_{yx_1\cdot x_2} = \frac{r_{yx_1} - r_{yx_2}\cdot r_{x_1x_2}}{\sqrt{(1 - r^2_{yx_2})(1 - r^2_{x_1x_2})}}$

- între $Y$ şi $X_2$, excluzând influenţa lui $X_1$: $r_{yx_2 \cdot x_1} = \dfrac{r_{yx_2} - r_{yx_1} \cdot r_{x_1x_2}}{\sqrt{(1 - r^2_{yx_1})(1 - r^2_{x_1x_2})}}$

Coeficienţii de corelaţie parţială, folosind estimaţiile variaţelor, se calculează după relaţiile:

- între $Y$ şi $X_1$, excluzând pe $X_2$: $\eta_{yx_1 \cdot x_2} = \sqrt{\dfrac{s^2_{yx_1x_2}}{s^2_{yx_2}}}$

- între $Y$ şi $X_2$, excluzând pe $X_1$: $\eta_{yx_2 \cdot x_1} = \sqrt{\dfrac{s^2_{yx_1x_2}}{s^2_{yx_1}}}$

## Determinaţia multiplă şi determinaţia parţială

Măsurarea influenţei factorilor, simultană şi separată, se stabileşte prin indicatorii determinaţiei multiple, respectiv, a determinaţiei parţiale.

**Raportul determinaţiei totale.** Acest indicator, la nivelul unui eşantion, se notează cu $D_{yx_1x_2\ldots}$ şi *arată influenţa simultană a variabilelor factoriale asupra variabilei rezultative.* Poate fi calculat prin ridicarea la pătrat fie a coeficientului de corelaţie multiplă $R$, fie a raportului de corelaţie multiplă $\eta_{y \cdot x_1x_2\ldots}$, după relaţiile:

$$D_{yx_1x_2} = R^2 = \frac{r^2_{yx_1} + r^2_{yx_2} - 2\, r_{yx_1} \cdot r_{yx_2} \cdot r_{x_1x_2}}{1 - r^2_{x_1x_2}}$$

$$D_{yx_1x_2\ldots} = \eta^2_{yx_1x_2\ldots} = 1 - \frac{\sum (y_i - y_{x_1x_2\ldots})^2}{\sum (y_i - \bar{y})^2}$$

sau:

$$D_{yx_1x_2\ldots} = \eta^2_{yx_1x_2\ldots} = \frac{a \sum y + b_1 \sum x_1 y + \ldots + b_k \sum x_k y - \dfrac{1}{n}(\sum y)^2}{\sum y^2 - \dfrac{1}{n}(\sum y)^2}$$

Raportul determinaţiei multiple variază între: $0 \le D_{yx_1x_2\ldots} \le 1$.

Pentru a facilita interpretarea rezultatelor, $D_{yx_1x_2\ldots}$ se poate exprima procentual *şi arată cât la sută din variaţia lui y depinde de variaţia simultană a variabilelor factoriale considerate.*

BUCUREŞTI UNIVERSITY LIBRARY

**Raportul determinației parțiale.** Acest indicator arată influența separată a factorilor. Se calculează fie prin ridicarea la pătrat a coeficienților corelației parțiale, fie prin descompunerea determinației totale.

Indicatorii determinației parțiale, pentru o regresie multiplă liniară de forma $y_{x_1x_2} = a + b_1 x_1 + b_2 x_2$, calculați pe baza corelației parțiale, sunt:

$$d_{yx_1 \cdot x_2} = r^2_{yx_1 \cdot x_2} \text{ și } d_{yx_2 \cdot x_1} = r^2_{yx_2 \cdot x_1}$$

Plecând de la determinația multiplă totală, pentru $Y$ în funcție de $X_1$ și $X_2$, și înlocuind $a$ cu expresia lui rezultată din prima ecuație a sistemului de ecuații normale în care s-au substituit $b_1$ și $b_2$:

$$D_{yx_1x_2} = \frac{a\sum y + b_1\sum x_1 y + b_2\sum x_2 y - \frac{1}{n}(\sum y)^2}{\sum y^2 - \frac{1}{n}(\sum y)^2}$$

$$a = \frac{\sum y - b_1\sum x_1 - b_2\sum x_2}{n} = \bar{y} - b_1\bar{x}_1 - b_2\bar{x}_2$$

se obține:

$$D_{yx_1x_2} = \frac{\bar{y}\sum y + b_1\bar{x}_1\sum y + b_2\bar{x}_2\sum y + b_1\sum x_1 y + b_2\sum x_2 y - \frac{1}{n}(\sum y)^2}{\sum y^2 - \frac{1}{n}(\sum y)^2} =$$

$$= \frac{b_1(\sum x_1 y - \frac{\sum x_1 \sum y}{n}) + b_2(\sum x_2 y - \frac{\sum x_2 \sum y}{n})}{\sum y^2 - \frac{1}{n}(\sum y)^2}$$

din care, indicatorii determinației parțiale se obțin prin divizarea determinației multiple totale:

$$D_{yx_1x_2} = \frac{b_1\sum x_1 y - \frac{\sum x_1 \sum y}{n}}{\sum y^2 - \frac{1}{n}(\sum y)^2} + \frac{b_2\sum x_2 y - \frac{\sum x_2 \sum y}{n}}{\sum y^2 - \frac{1}{n}(\sum y)^2}$$

sau $$D_{yx_1x_2} = \frac{b_1(n\sum x_1 y - \sum x_1 \sum y)}{n\sum y^2 - (\sum y)^2} + \frac{b_2(n\sum x_2 y - \sum x_2 \sum y)}{n\sum y^2 - (\sum y)^2}$$

Primul termen al membrului doi al relației reprezintă **determinația parțială între $Y$ și $X_1$**, iar al doilea, **determinația parțială între $Y$ și $X_2$.**

Prin urmare,

$$D_{yx_1x_2} = d_{yx_2\cdot x_1} + d_{yx_1\cdot x_2}$$

## Coeficientul de determinaţie ajustat

Pentru un model liniar multiplu cu două variabile independente, coeficientul de determinaţie multiplă se estimează pe baza relaţiei de descompunere a variaţiei totale:

$$\sum_i (y_i - \bar{y})^2 = \sum_i (y_{xi} - \bar{y})^2 + \sum_i (y_i - y_{xi})^2$$

sau, în SPSS, *TSS = ESS + RSS (TSS – estimaţia variaţiei totale, ESS – estimaţia variaţiei explicate, RSS – estimaţia variaţiei reziduale).*

Relaţia estimaţiei coeficientului determinaţiei multiple este:

$$R^2 = \frac{ESS}{TSS} = \frac{b_1 \sum_i (y_i - \bar{y})(x_{1i} - \bar{x}_1) + b_2 \sum_i (y_i - \bar{y})(x_{2i} - \bar{x}_2)}{\sum_i (y_i - \bar{y})^2}$$

sau $$R^2 = 1 - \frac{RSS}{TSS} = 1 - \frac{\sum_i e_i^2}{\sum_i (y_i - \bar{y})^2}$$

*Observaţie*

Coeficientul de determinaţie multiplă nu ţine seama de numărul de grade de libertate sau de numărul de parametri care apar în model.

Pentru a înlătura neajunsul precizat mai sus, se determină coeficientul de determinaţie ajustat prin relaţia:

$$R^2_{adj} = 1 - \frac{\frac{\sum_i e_i^2}{n-k}}{\frac{\sum_i (y_i - \bar{y})^2}{n-1}} = 1 - \frac{s_\varepsilon^2}{s_y^2}$$

Scris altfel,

$$R^2_{adj} = 1 - (1 - R^2)\frac{n-1}{n-k}.$$

BCU IASI CENTRAL UNIVERSITY LIBRARY

*Observaţie*

$R^2_{adj} < R^2$, pentru $k>1$.

$R^2_{adj}$ poate fi şi negativ, când coeficientul de determinaţie este aproape de zero.

## Testarea modelului de regresie – Testul F

Testarea modelului de regresie liniară multiplă se realizează cu ajutorul testului *F*, conform următorului demers:

- Ipoteze: $H_0 : \alpha = \beta_1 = ... = \beta_p = 0$

$H_1$ : nu toţi coeficienţii sunt simultan zero;

- Statistica Fisher: $F = \frac{\hat{R}^2}{1-\hat{R}^2}\frac{n-k}{k-1} \sim F(k-1, n-k)$.

*Observaţie*

Pentru două variabile independente, $k = 3$, statistica Fisher se va scrie:

$$F = \frac{\hat{R}^2}{1-\hat{R}^2}\frac{n-k}{k-1} = \frac{\hat{R}^2}{1-\hat{R}^2}\cdot\frac{n-3}{2} \sim F(2, n-3).$$

- Valoarea teoretică $F_{\alpha(1,n-2)}$ din tabela Fisher;

- valoarea calculată la nivelul eşantionului $F_{calc.}$:

$$F_{calc.} = \frac{ESS}{RSS}\frac{n-k}{k-1} = \frac{R^2}{1-R^2}\cdot\frac{n-k}{k-1}.$$

Pentru $k = 3$, relaţia este:

$$F_{calc.} = \frac{R^2}{1-R^2}\cdot\frac{n-3}{2}.$$

- regula de decizie: dacă $F_{calc.} > F_{\alpha(2,n-3)}$, se respinge ipoteza $H_0$ cu o probabilitate egală cu $(1-\alpha)$.

În SPSS, dacă $sigF<\alpha$ se respinge $H_0$.

## Testarea influenţei marginale a unei variabile independente

În cazul în care se doreşte testarea influenţei marginale a unei variabile noi introduse în model, se foloseşte un test Fisher, cu o statistică dată prin relaţia:

BCU IAŞI - BIBLIOTECA UNIVERSITARĂ

$$F = \frac{ESS_{new} - ESS_{old}}{RSS_{new}} \frac{n - k_{new}}{k_{new} - 1} = \frac{R^2_{new} - R^2_{old}}{1 - R^2_{new}} \frac{n - k_{new}}{k_{new} - 1},$$

unde „old" specifică indicatorul înainte de introducerea variabilei independente în model, iar „new" indicatorul după introducerea variabilei în model.

În acest caz, ipotezele sunt următoarele:

$H_0$ : variabila introdusă în model nu are o influență semnificativă;

$H_1$ : caz contrar, variabila are o influență semnificativă asupra variabilei dependente.

### 3.1.5. Regresia multiplă în SPSS

Pentru a găsi cea mai bună combinație de variabile independente care explică variația variabilei dependente, într-un model de regresie, SPSS oferă mai multe metode: **Forward**, **Backward**, **Stepwise**. Prin aceste metode se pot selecta variabile care explică cel mai bine variația variabilei dependente. Aplicarea lor presupune introducerea și eliminarea variabilelor independente în model în funcție de gradul de semnificație a legăturii lor cu variabila dependentă, până când nici o variabilă nu mai poate fi introdusă sau eliminată din ecuația de regresie.

**1. Forward (Introducerea pas cu pas).** Prin acest procedeu, variabilele independente sunt introduse în model una câte una (pas cu pas), în ordinea importanței lor. În pasul întâi, este introdusă variabila care este cel mai puternic corelată, pozitiv sau negativ, cu variabila dependentă. În pasul doi (și următorii) se introduc variabile mai puțin corelate. La fiecare pas, este testată ipoteza de nul asupra coeficientului de regresie a variabilei introduse, adică se testează dacă coeficientul de regresie corespunzător este zero. Este folosită statistica **test t** (respectiv, statistica **F** care este pătratul statisticii **t**). Pașii se opresc când un prag de semnificație stabilit pentru **F** nu mai este atins.

**2. Backward (Eliminarea pas cu pas).** Acest procedeu este cel mai des folosit în practică. Începe cu toate variabilele considerate în model și la fiecare pas se elimină cel mai slab predictor (variabilă independentă). Cel mai slab predictor este definit de variabila independentă cel mai puțin importantă, adică variabila care determină cea mai mică reducere a statisticii Fisher, **F**. Variabilele sunt eliminate până când un prag de semnificație stabilit pentru **F** nu mai este atins.

3. **Stepwise (Selecția pas cu pas).** Acest procedeu începe la fel ca Forward, dar la fiecare pas, testează variabilele existente deja în model, pentru a le elimina. Aceasta este metoda cea mai folosită, în special, când există corelații între variabilele independente. De exemplu, introducerea celei de a patra variabile poate diminua importanța unei variabile deja introduse și ca urmare, aceasta este eliminată din model (în Forward aceasta rămâne în model).

**Exemplu**

Pentru realizarea în SPSS a unei analize de regresie multiplă vom considera datele din **reg_pib_inv_cs_pocup.sav**, referitoare la regiunile României, în anul 2000 și procedeul **Backward.**

Pașii demersului sunt cei prezentați în cazul unui model de regresie liniar simplă, cu elementele specifice unui model de regresie multiplă. Astfel, în fereastra **Linear Regression** selectăm:

- în zona **Dependent**: câștigul salarial nominal net (**cs**);
- în zona **Independent(s)**: produsul intern brut regional (**pib**), investițiile în 2000 ( **inv**) și populația ocupată în 2000 (**pocup**);
- în zona **Method**: se alege metoda **Backward**;
- în zona **Case Labels** : regiunile României (**reg**);
- în fereastra **Linear Regression: Statistics**, deschisă prin butonul de comandă **Statistics**, se activează casetele de validare **Descriptives** și **Collinearty diagnostics.**

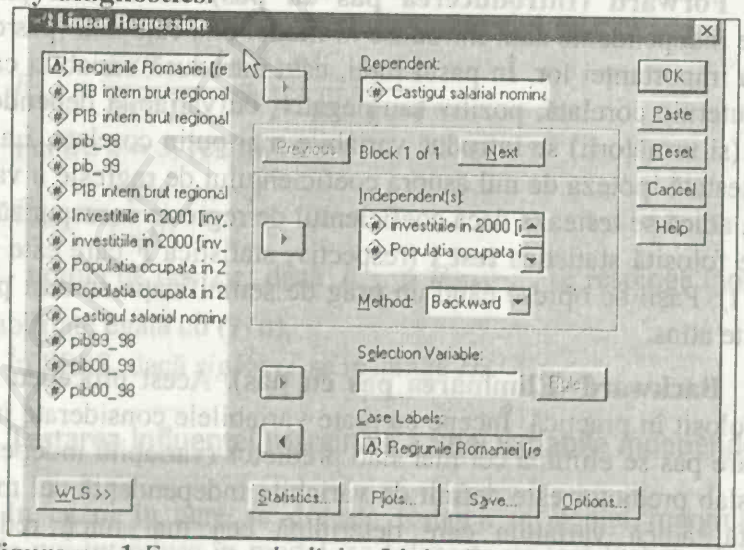



**Figura nr. 1** *Fereastra de dialog* **Liniar Regression**, *metoda* **Backward**

În tabelul **Correlations** se afişează coeficienţii de corelaţie Pearson (**Pearson Correlation**), valoarea semnificaţiei (**Sig.**) pentru fiecare coeficient de corelaţie şi numărul cazurilor considerate în studiu (N).

Pentru exemplul dat, sunt prezentate corelaţiile simple ale fiecărei variabile independente (predictor) cu variabila dependentă **cs** – câştigul salarial nominal net (vezi matricea corelaţiilor din Tabelul 1).

**Tabelul nr. 1.** *Matricea corelaţiilor parţiale*

Correlations

| | | Castigul salarial nominal net in anul 2000 | PIB intern brut regional pe locuitor in anul 2000 (lei) | Investitiile in 2000 | Populatia ocupata in 2000 (mii persoane) |
|---|---|---|---|---|---|
| Pearson Correlation | Castigul salarial nominal net in anul 2000 | 1.000 | .001 | .877 | -.555 |
| | PIB intern brut regional pe locuitor in anul 2000 (lei) | .001 | 1.000 | .157 | -.710 |
| | Investitiile in 2000 | .877 | .157 | 1.000 | -.737 |
| | Populatia ocupata in 2000 (mii persoane) | -.555 | -.710 | -.737 | 1.000 |
| Sig. (1-tailed) | Castigul salarial nominal net in anul 2000 | . | .500 | .005 | .098 |
| | PIB intern brut regional pe locuitor in anul 2000 (lei) | .500 | . | .368 | .037 |
| | Investitiile in 2000 | .005 | .368 | . | .029 |
| | Populatia ocupata in 2000 (mii persoane) | .098 | .037 | .029 | . |
| N | Castigul salarial nominal net in anul 2000 | 7 | 7 | 7 | 7 |
| | PIB intern brut regional pe locuitor in anul 2000 (lei) | 7 | 7 | 7 | 7 |
| | Investitiile in 2000 | 7 | 7 | 7 | 7 |
| | Populatia ocupata in 2000 (mii persoane) | 7 | 7 | 7 | 7 |

Se observă că valoarea coeficienţilor de corelaţie de pe diagonală este egală cu unu, deoarece fiecare variabilă este corelată perfect cu ea însăşi. Se constată că legătura cea mai semnificativă este între câştigul salarial nominal net şi investiţii. Între variabila dependentă – **cs** şi variabila independentă – **inv** există o legătură directă, puternică. Valoarea coeficientului de corelaţie este egală cu 0,877, cu o valoare **Sig.** mai mică decât 0,05.

Tabelul **Variable Entered/Removed** furnizează o prezentare a rezultatelor eliminării pas cu pas a variabilelor (vezi Tabelul 2).

SPSS elaborează, la început, un model cu toate variabilele independente, folosind metoda **Enter**, apoi, în fiecare pas creează un model, eliminând variabila care are cea mai redusă contribuţie.

**Tabelul nr. 2.** *Variabilele introduse în model şi variabilele eliminate pas cu pas*

**Variables Entered/Removed[b]**

| Model | Variables Entered | Variables Removed | Method |
|---|---|---|---|
| 1 | Populatia ocupata in 2000 (mii persoane), PIB intern brut regional pe locuitor in anul 2000 (lei), investitiile in 2000[a] | . | Enter |
| 2 | . | Populatia ocupata in 2000 (mii persoane) | Backward (criterion: Probability of F-to-remove >= .100). |
| 3 | . | PIB intern brut regional pe locuitor in anul 2000 (lei) | Backward (criterion: Probability of F-to-remove >= .100). |

a. All requested variables entered.

b. Dependent Variable: Castigul salarial nominal net in anul 2000

În exemplul considerat, sunt eliminate, pe rând, în ordinea celei mai slabe influenţe asupra câştigului salarial nominal net, variabila populaţie ocupată şi variabila produs intern brut pe locuitor.

Tabelul **Model Summary** prezintă pentru fiecare model de regresie valoarea raportului de corelaţie (**R**), valoarea coeficientului de determinaţie ($\mathbf{R^2}$) şi eroarea standard. Valoarea $\mathbf{R^2}$ creşte pe măsură ce se introduc mai multe variabile în model. Includerea de variabile irelevante duce, de asemenea, la creşterea erorii standard.

BCU IASI

**Tabelul nr. 3.** *Model Summary, cazul regresiei multiple*

Model Summary[d]

| Model | R | R Square | Adjusted R Square | Std. Error of the Estimate | Change Statistics | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | | | | R Square Change | F Change | df1 | df2 | Sig. F Change |
| 1 | .888[a] | .789 | .578 | 130351.7763 | .789 | 3.745 | 3 | 3 | .153 |
| 2 | .888[b] | .789 | .683 | 113053.5944 | -.001 | .009 | 1 | 5 | .931 |
| 3 | .877[c] | .769 | .723 | 105627.4703 | -.019 | .365 | 1 | 6 | .578 |

a. Predictors: (Constant), Populatia ocupata in 2000 (mii persoane), PIB intern brut regional pe locuitor in anul 2000 (lei), Investitiile in 2000

b. Predictors: (Constant), PIB intern brut regional pe locuitor in anul 2000 (lei), Investitiile in 2000

c. Predictors: (Constant), Investitiile in 2000

d. Dependent Variable: Castigul salarial nominal net in anul 2000

În exemplul dat, valoarea **R**, valoarea $\mathbf{R^2}$ ajustat şi eroarea standard arată că cel mai bun predictor (variabila independentă care estimează cel mai bine variabila dependentă) este variabila investiţii.

Aceeaşi concluzie se poate trage considerând rezultatele din tabelul ANOVA (vezi Tabelul 3). Dacă valoarea semnificaţiei statisticii **F** este mică (**Sig.** este mai mică decât 0,05), atunci variabilele independente explică variaţia variabilei dependente. Cea mai mică valoare **Sig.** corespunde modelului care explică variaţia câştigului salarial nominal net în funcţie de investiţii.

În tabelul **coeficienţilor de regresie**, în prima parte, apar coeficienţii de regresie, erorile standard, valoarea statisticii test t pentru fiecare coeficient, precum şi valoarea **Sig.** În cazul unei regresii multiple apare, în plus, faţă de cazul unei corelaţii simple statisticile de coliniaritate (*collinearity statistics*), toleranţa (*tolerance*) şi factorul de inflaţie a varianţei (*variance inflation factor* – VIF).

**Tabelul nr. 4.** *ANOVA*

ANOVA[d]

| Model | | Sum of Squares | df | Mean Square | F | Sig. |
|---|---|---|---|---|---|---|
| 1 | Regression | 1.91E+11 | 3 | 6.363E+10 | 3.745 | .153[a] |
| | Residual | 5.10E+10 | 3 | 1.699E+10 | | |
| | Total | 2.42E+11 | 6 | | | |
| 2 | Regression | 1.91E+11 | 2 | 9.536E+10 | 7.461 | .045[b] |
| | Residual | 5.11E+10 | 4 | 1.278E+10 | | |
| | Total | 2.42E+11 | 6 | | | |
| 3 | Regression | 1.86E+11 | 1 | 1.861E+11 | 16.677 | .010[c] |
| | Residual | 5.58E+10 | 5 | 1.116E+10 | | |
| | Total | 2.42E+11 | 6 | | | |

a. Predictors: (Constant), Populatia ocupata in 2000 (mii persoane), PIB intern brut regional pe locuitor in anul 2000 (lei), investitiile in 2000

b. Predictors: (Constant), PIB intern brut regional pe locuitor in anul 2000 (lei), investitiile in 2000

c. Predictors: (Constant), investitiile in 2000

d. Dependent Variable: Castigul salarial nominal net in anul 2000

**Coliniaritatea** exprimă existența unei corelații puternice între variabilele independente. În astfel de situații, se calculează statisticile toleranței, considerând numai variabilele independente, variabila dependentă este exclusă din model.

**Toleranța** fiecărei variabile $X_i$ se calculează după relația:

Toleranța $= 1 - R_i^2$,

unde:

$R_i^2$ este pătratul coeficientului de corelație multiplă a variabilei $X_i$ cu toate celelalte variabile independente.

**VIF** este reciproca toleranței.

Toleranța poate lua valori de la 0 la 1. Cu cât valoarea toleranței este mai mică, mai apropiată de zero, cu atât variabila independentă $X_i$ este explicată printr-o combinație liniară a celorlalte variabile independente. Ca urmare, explicarea variabilei dependente prin această variabilă poate fi considerată că are prea puțină acuratețe.

**Tabelul nr. 5.** *Coeficienții de regresie*

Coefficients[a]

| Model | | Unstandardized Coefficients | | Standardized Coefficients | t | Sig. | 95% Confidence Interval for B | | Collinearity Statistics | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | B | Std. Error | Beta | | | Lower Bound | Upper Bound | Tolerance | VIF |
| 1 | (Constant) | 2191707 | 2170142 | | 1.010 | .387 | -4714653.203 | 9098066.708 | | |
| | PIB intern brut regional pe locuitor in anul 2000 (lei) | -4.01E-03 | .026 | -.091 | -.155 | .887 | -.097 | .079 | .202 | 4.954 |
| | investitiile in 2000 | 42.296 | 27.310 | .951 | 1.549 | .219 | -44.615 | 129.208 | .186 | 5.367 |
| | Populatia ocupata in 2000 (mil persoane) | 81.670 | 870.088 | .081 | .094 | .931 | -2687.337 | 2850.677 | .095 | 10.566 |
| 2 | (Constant) | 2391840 | 350602.0 | | 6.822 | .002 | 1418412.950 | 3365267.294 | | |
| | PIB intern brut regional pe locuitor in anul 2000 (lei) | -6.18E-03 | .010 | -.141 | -.604 | .578 | -.035 | .022 | .975 | 1.025 |
| | investitiile in 2000 | 39.991 | 10.352 | .899 | 3.863 | .018 | 11.248 | 68.733 | .975 | 1.025 |
| 3 | (Constant) | 2210270 | 168505.7 | | 13.117 | .000 | 1777112.723 | 2643428.213 | | |
| | investitiile in 2000 | 39.010 | 9.552 | .877 | 4.084 | .010 | 14.454 | 63.565 | 1.000 | 1.000 |

a. Dependent Variable: Castigul salarial nominal net in anul 2000

**Diagnosticul coliniarității** presupune analiza rezultatelor din tabelul **Collinearity Diagnostics** (vezi Tabelul 5).

**Tabelul nr. 6.** *Diagnosticul coliniarității*

Collinearity Diagnostics[a]

| Model | Dimension | Eigenvalue | Condition Index | Variance Proportions | | | |
|---|---|---|---|---|---|---|---|
| | | | | (Constant) | PIB intern brut regional pe locuitor in anul 2000 (lei) | investitiile in 2000 | Populatia ocupata in 2000 (mil persoane) |
| 1 | 1 | 3.906 | 1.000 | .00 | .00 | .00 | .00 |
| | 2 | 7.156E-02 | 7.388 | .00 | .00 | .08 | .02 |
| | 3 | 2.237E-02 | 13.214 | .00 | .10 | .10 | .02 |
| | 4 | 3.752E-04 | 102.030 | 1.00 | .90 | .83 | .97 |
| 2 | 1 | 2.954 | 1.000 | .00 | .00 | .01 | |
| | 2 | 3.752E-02 | 8.873 | .05 | .09 | .97 | |
| | 3 | 8.423E-03 | 18.727 | .95 | .91 | .02 | |
| 3 | 1 | 1.972 | 1.000 | .01 | | .01 | |
| | 2 | 2.847E-02 | 8.321 | .99 | | .99 | |

a. Dependent Variable: Castigul salarial nominal net in anul 2000

**Eigenvalue** dă o indicație asupra numărului de legături care există între variabilele independente. Când mai multe **eigenvalues** sunt apropiate de zero, variabilele sunt puternic intercorelate.

**Indicii de condiție** se calculează ca rădăcină pătrată din raportul dintre valoarea eigenvalue cea mai mare și valoarea eigenvalue a fiecărei dimensiuni. Un indice mai mare decât 15 arată că există o posibilă problemă de coliniaritate, iar o valoare mai mare ca 30 indică probleme grave de coliniaritate. Aceste situații le întâlnim în exemplul considerat, pentru modelul 1, indicele corespunzător dimensiunii 4 (variabila populația ocupată) are valoarea de 102,030, respectiv, pentru modelul 2, indicele corespunzător dimensiunii 3 (variabila PIB) are valoarea de 18,721 (vezi Tabelul 6).

**Proporția varianței** evidențiază contribuția fiecărei variabile la varianță. Variabilele care au valori mari pentru acest indicator arată probleme de coliniaritate. În exemplul dat, variabilele cu probleme de coliniaritate și care influențează substanțial varianța sunt:

- populația ocupată, cu o proporție de 0,97;
- PIB regional, cu o proporție de 0,91.

Tabelul **Excluded Variables** prezintă informații despre variabilele care sunt excluse la fiecare pas (vezi Tabelul 7).

**Tabelul nr. 7.** *Variabile excluse*

Excluded Variables[c]

| Model | | Beta In | t | Sig. | Partial Correlation | Collinearity Statistics | | |
|---|---|---|---|---|---|---|---|---|
| | | | | | | Tolerance | VIF | Minimum Tolerance |
| 2 | Populatia ocupata in 2000 (mil persoane) | .081[a] | .094 | .931 | .054 | 9.464E-02 | 10.566 | 9.464E-02 |
| 3 | Populatia ocupata in 2000 (mil persoane) | .200[b] | .586 | .590 | .281 | .457 | 2.187 | .457 |
| | PIB intern brut regional pe locuitor in anul 2000 (lei) | -.141[b] | -.604 | .578 | -.289 | .975 | 1.025 | .975 |

a. Predictors in the Model: (Constant), PIB intern brut regional pe locuitor in anul 2000 (lei), investitiile in 2000

b. Predictors in the Model: (Constant), investitiile in 2000

c. Dependent Variable: Castigul salarial nominal net in anul 2000

**Beta in** este coeficientul de regresie care ar rezulta dacă în pasul următor s-ar păstra în model variabila exclusă.

**Statistica test t** și **valoarea Sig.** sunt folosite pentru testarea ipotezei de nul cu privire la coeficienții de regresie, adică a ipotezei că între variabila dependentă și variabila independentă nu există o legătură semnificativă.

În exemplul considerat, se constată valori **Sig.** foarte mari (comparativ cu 0.05), ceea ce nu ne permite să respingem ipoteza de nul,

adică a inexistenței unei legături semnificative între variabila dependentă – câștigul salarial – și variabilele independente – populația ocupată și PIB regional pe locuitor, la nivelul anului 2000, în România.

Se observă, de asemenea, valori mici pentru toleranță și valori mari pentru VIF, ceea ce denotă existența multicoliniarității, care determină o varianță mare a coeficientului de regresie, și, ca urmare, o instabilitate a estimației.

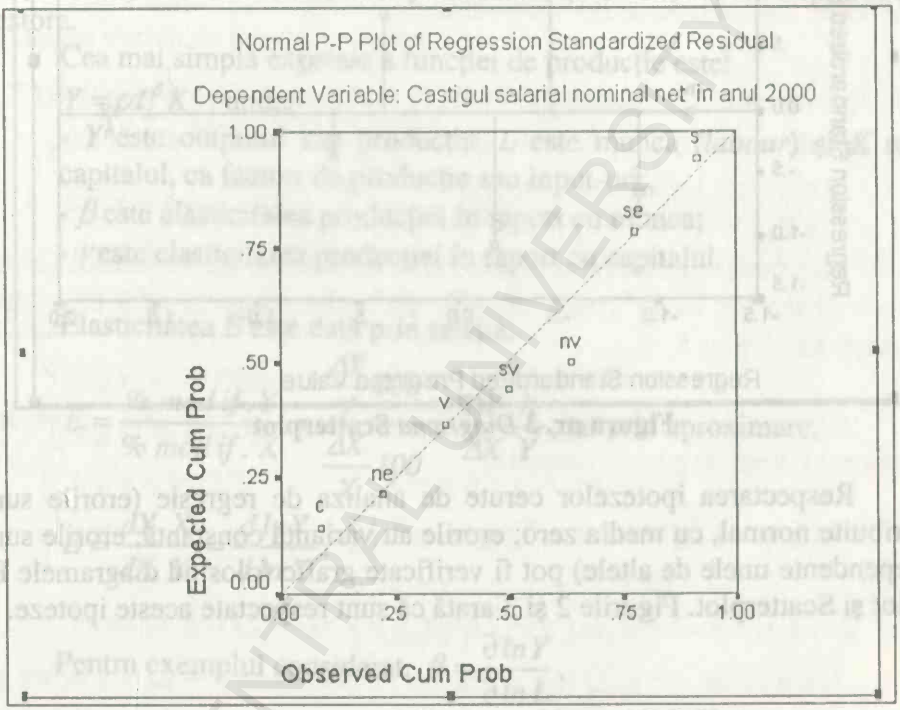



**Figura nr. 2** *Diagrama* **Normal P-P Plot**

BCU IASI CENTRAL UNIVERSITY LIBRARY

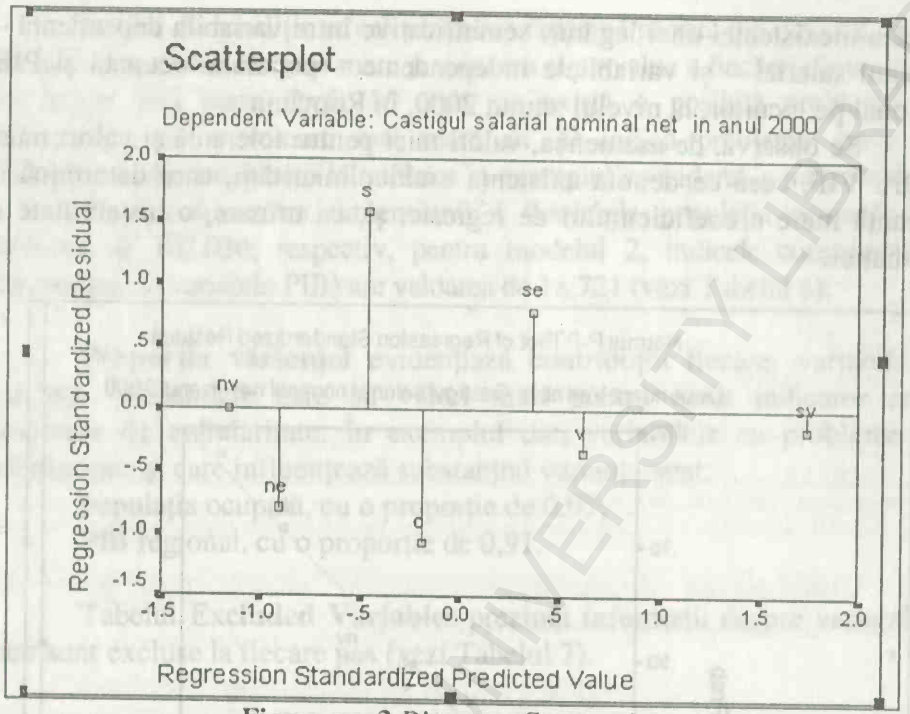



**Figura nr. 3** *Diagrama* **Scatterplot**

Respectarea ipotezelor cerute de analiza de regresie (erorile sunt distribuite normal, cu media zero; erorile au varianță constantă; erorile sunt independente unele de altele) pot fi verificate grafic folosind diagramele P-P Plot și Scatterplot. Figurile 2 și 3 arată că sunt respectate aceste ipoteze.

## 3.2. Modele neliniare multiple

Complexitatea fenomenelor economice a condus la elaborarea unor modele multiple neliniare. Ca și în cazul modelelor simple, modelele multiple sunt de mai multe tipuri, dintre care prezentăm modelele liniarizabile și modelele polinomiale.

BCU IOAN SLAVICI UNIVERSITY LIBRARY

### 3.2.1. Prezentarea problemei. Exemple din economie

Poate cel mai cunoscut model din această categorie este funcția de producție, numită și funcția Cobb-Douglas. Această funcție exprimă relația dintre output și input în cazul unei firme sau la nivelul unei economii naționale. Există o multitudine de expresii ale funcției de producție, în funcție de numărul de factori luați în calcul și de modul de exprimare a acestora.

Cea mai simplă expresie a funcției de producție este:

$Y = \alpha L^{\beta} K^{\gamma}$, unde:

- $Y$ este outputul sau producția, $L$ este munca (*labour*) și $K$ este capitalul, ca factori de producție sau input-uri;
- $\beta$ este elasticitatea producției în raport cu munca;
- $\gamma$ este elasticitatea producției în raport cu capitalul.

Elasticitatea $E$ este dată prin relația:

$$E = \frac{\%\ mod\ if.\ Y}{\%\ mod\ if.\ X} = \frac{\frac{\Delta Y}{Y}100}{\frac{\Delta X}{X}100} = \frac{\Delta Y}{\Delta X}\frac{X}{Y}, \text{ iar prin aproximare,}$$

$$E = \frac{dY}{dX}\frac{X}{Y} = \frac{d\ln Y}{d\ln X}.$$

Pentru exemplul considerat, $\beta = \frac{\partial \ln Y}{\partial \ln L}$.

### 3.2.2. Modele liniarizabile

Între modelele liniarizabile se află modelele de tip exponențial și putere.

**Modelul putere**

Ecuația estimată a modelului regresiei multiple neliniare, pentru o variabilă dependentă $Y$ și variabilele factoriale $X_1, X_2, ..., X_p$, de tipul **funcțiilor putere**, ia forma:

$$Y_{x_1 x_2 \ldots} = a\, X_1^{b_1} \cdot X_2^{b_2} \ldots X_p^{b_p}$$

Pentru facilitatea calculelor, modelul se liniarizează, devenind:

$$\log Y_{x_1x_2\ldots} = \log a + b_1 \log X_1 + b_2 \log X_2 + \ldots + b_p \log X_p$$

Obţinerea parametrilor presupune rezolvarea sistemului corespunzător de ecuaţii normale rezultat din aplicarea metodei celor mai mici pătrate pentru expresia anterioară.

Un model de corelaţie bifactorială, utilizat mult în modelarea creşterii economice, este funcţia de **tip Cobb-Douglas**:

$$Y_{x_1x_2} = a\, X_1^{b_1} \cdot X_2^{b_2},$$

care exprimă corelarea produsului final ($Y$) cu fondurile fixe productive ($X_1$) şi cu forţa de muncă ($X_2$); $b_1$ şi $b_2$ reprezintă coeficienţii de elasticitate.

În expresie logaritmică, modelul de mai sus devine:

$$\log Y_{x_1x_2} = \log a + b_1 \log X_1 + b_2 \log X_2$$

Prin rezolvarea sistemului de ecuaţii normale (aplicând metoda Gauss, metoda abaterilor de la medie, metoda determinanţilor) se află estimaţiile parametrilor: $\log a$; $b_1$; $b_2$.

Estimaţiile parametrilor ecuaţiei de regresie se află rezolvând următorul sistem de ecuaţii normale:

$$n \log a + b_1 \sum \log x_1 + b_2 \sum \log x_2 = \sum \log y$$

$$\log a \sum \log x_1 + b_1 \sum (\log x_1)^2 + b_2 \sum (\log x_1 \log x_2) = \sum (\log x_1 \log y)$$

$$\log a \sum \log x_2 + b_1 \sum (\log x_1 \log x_2) + b_2 \sum (\log x_2)^2 = \sum (\log x_2 \log y)$$

**Exemplu.** În vederea studierii legăturii dintre producţia industrială ($Y$), personalul muncitor ($X_1$) şi fondurile fixe din industrie ($X_2$), considerăm ponderile exprimate procentual ale acestor variabile pe judeţe faţă de total ţară, redate în tabelul 8.

Se cere:

1. Să se estimeze parametrii ecuaţiei de regresie multiplă;
2. Să se determine intensitatea corelaţiei multiple;
3. Să se calculeze raportul determinaţiei multiple şi coeficienţii determinaţiei parţiale.

BCU IASI / CENTRAL UNIVERSITY LIBRARY

**Rezolvare**

**1.** În teoria şi practica modelării creşterii economice se consideră că între variabilele admise există o legătură exprimată printr-un model de corelaţie bifactorială redat printr-o funcţie de tip Cobb-Douglas:

$$Y_{x_1x_2} = a\, X_1^{b_1} \cdot X_2^{b_2}$$

care prin logaritmare devine:

$$\log Y = \log a + b_1 \log X_1 + b_2 \log X_2$$

Estimaţiile parametrilor ecuaţiei de regresie se află rezolvând următorul sistem de ecuaţii normale:

$$n \log a + b_1 \sum \log x_1 + b_2 \sum \log x_2 = \sum \log y$$

$$\log a \sum \log x_1 + b_1 \sum (\log x_1)^2 + b_2 \sum (\log x_1 \log x_2) = \sum (\log x_1 \log y)$$

$$\log a \sum \log x_2 + b_1 \sum (\log x_1 \log x_2) + b_2 \sum (\log x_2)^2 = \sum (\log x_2 \log y)$$

Calcule:

$$\sum \log x_1 = 10{,}7989 \qquad \sum (\log x_1)^2 = 5{,}29089$$

$$\sum \log x_2 = 10{,}9084 \qquad \sum (\log x_2)^2 = 5{,}0203$$

$$\sum \log y = 9{,}82793 \qquad \sum (\log x_1 \log x_2) = 4{,}96492$$

$$n = 40 \qquad \sum (\log x_1 \log y) = 5{,}16665$$

$$\sum (\log x_2 \log y) = 5{,}2936$$

$$\sum (\log y)^2 = 5{,}49318$$

**Tabelul nr. 8.** *Ponderea producţiei industriale (Y), a personalului muncitor ( $X_1$ ) şi a fondurilor fixe ( $X_2$ ) pe judeţe faţă de total ţară (în %)*

| Judeţul | $y$ | $x_1$ | $x_2$ | Judeţul | $y$ | $x_1$ | $x_2$ |
|---|---|---|---|---|---|---|---|
| 1 | 2 | 3 | 3 | 1 | 2 | 3 | 4 |
| 1. Alba | 1,7 | 2,2 | 1,5 | 21. Harghita | 1,1 | 1,8 | 1,0 |
| 2. Arad | 1,8 | 2,3 | 1,5 | 22. Hunedoara | 3,5 | 3,2 | 5,1 |
| 3. Argeş | 4,4 | 3,3 | 4,8 | 23. Ialomiţa | 0,7 | 0,5 | 0,8 |
| 4. Bacău | 3,6 | 2,9 | 4,2 | 24. Iaşi | 2,8 | 2,7 | 2,7 |
| 5. Bihor | 2,3 | 2,9 | 1,9 | 25. Maramureş | 1,9 | 2,4 | 1,6 |
| 6. Bistriţa-Năsăud | 0,7 | 1,0 | 0,8 | 26. Mehedinţi | 1,0 | 0,9 | 1,2 |
| 7. Botoşani | 0,8 | 1,1 | 0,7 | 27. Mureş | 2,8 | 3,0 | 2,6 |
| | | | | 28. Neamţ | 2,3 | 2,5 | 2,9 |

BIBLIOTECA CENTRALĂ UNIVERSITY LIBRARY

| Județul | $y$ | $x_1$ | $x_2$ | Județul | $y$ | $x_1$ | $x_2$ |
|---|---|---|---|---|---|---|---|
| 1 | 2 | 3 | 3 | 1 | 2 | 3 | 4 |
| 8. Brașov | 5,8 | 5,2 | 5,1 | 29. Olt | 2,2 | 1,5 | 2,4 |
| 9. Brăila | 1,7 | 1,6 | 2,1 | 30. Prahova | 7,1 | 5,2 | 5,4 |
| 10. Buzău | 1,7 | 1,9 | 1,6 | 31. Satu-Mare | 1,2 | 1,7 | 0,7 |
| 11. Caraș-Severin | 2,0 | 2,1 | 2,2 | 32. Sălaj | 0,9 | 0,9 | 0,9 |
| 12. Călărași | 0,7 | 0,8 | 0,7 | 33. Sibiu | 3,3 | 3,4 | 1,8 |
| 13. Cluj | 3,1 | 3,9 | 2,9 | 34. Suceava | 1,6 | 2,4 | 1,6 |
| 14. Constanța | 2,3 | 2,0 | 2,1 | 35. Teleorman | 1,2 | 1,3 | 1,8 |
| 15. Covasna | 0,9 | 1,1 | 0,9 | 36. Timiș | 3,5 | 3,9 | 2,5 |
| 16. Dâmbovița | 2,3 | 2,6 | 3,3 | 37. Tulcea | 0,9 | 0,9 | 1,7 |
| 17. Dolj | 2,5 | 2,5 | 2,9 | 38. Vaslui | 1,1 | 1,4 | 1,2 |
| 18. Galați | 5,0 | 2,5 | 4,6 | 39. Vâlcea | 1,3 | 1,3 | 2,7 |
| 19. Giurgiu | 0,5 | 0,6 | 0,9 | 40. Vrancea | 0,8 | 1,0 | 0,7 |
| 20. Gorj | 1,4 | 1,8 | 4,5 | | | | |

*Sursa: Anuarul Statistic al României, 1984, pp. 52-53.*

Înlocuind în sistem valorile sumelor variabilelor, obținem:

$$40 \log a + 10{,}7985\, b + 10{,}9084\, b = 9{,}82793$$
$$10{,}7989 \log a + 5{,}29089\, b_1 + 4{,}96492\, b_2 = 5{,}16665$$
$$10{,}9084 \log a + 4{,}96492\, b_1 + 5{,}0203\, b_2 = 5{,}2936$$

Prin rezolvarea sistemului obținem următoarele valori pentru estimațiile parametrilor ecuației regresiei multiple:

$$\log a = -0{,}056676$$
$$b_1 = 0{,}728269$$
$$b_2 = 0{,}387821$$

Ecuația de regresie multiplă va fi:

$$\log Y = -0{,}056676 + 0{,}728269 \log X_1 + 0{,}387821 \log X_2$$

**2.** Intensitatea corelației dintre variabilele luate în studiu se calculează cu ajutorul raportului corelației multiple, după relația:

$$\eta_{y_x} = \sqrt{\frac{\log a \sum \log y + b_1 \sum (\log x_1 \log y) + b_2 \sum (\log x_2 \log y) - \frac{1}{n}(\sum \log y)^2}{\sum (\log y)^2 - \frac{1}{n}(\sum \log y)^2}} =$$

$$= \sqrt{\frac{-0{,}056676 \cdot 9{,}82793 + 0{,}728269 \cdot 5{,}16665 + 0{,}387821 \cdot 5{,}2936 - \frac{1}{40}(9{,}82793)^2}{5{,}49318 - \frac{1}{40}(9{,}82793)^2}} = 0{,}9612$$

Valoarea raportului de corelaţie arată o legătură foarte strânsă între variabilele admise.

**Coeficienţii corelaţiei parţiale.**

Pentru calcularea lor s-au determinat coeficienţii de corelaţie simplă liniară ale căror valori sunt:

$$r_{\log y \log x_1} = \frac{n \sum (\log x_1 \log y) - \sum \log x_1 \sum \log y}{\sqrt{[n \sum (\log x_1)^2 - (\sum \log x_1)^2][n \sum (\log y)^2 - (\sum \log y)^2]}} =$$

$$= \frac{40 \cdot 5{,}16665 - 10{,}7989 \cdot 9{,}82793}{\sqrt{[40 \cdot 5{,}29089 - (10{,}7989)^2][40 \cdot 5{,}49318 - (9{,}82793)^2]}} = 0{,}929426$$

respectiv

$$r_{\log y \log x_2} = 0{,}867890 \quad r_{\log x_1 \log x_2} = 0{,}763632 .$$

Calculul coeficienţilor corelaţiei parţiale pe baza coeficienţilor de corelaţie simplă s-a efectuat după relaţiile:

$$r_{\log y \log x_1} = \frac{r_{\log y \log x_1} - r_{\log y \log x_2} \cdot r_{\log x_1 \log x_2}}{\sqrt{(1 - r^2_{\log y \log x_1})(1 - r^2_{\log x_1 \log x_2})}} = 0{,}831466$$

$$r_{\log y \log x_2} = \frac{r_{\log y \log x_2} - r_{\log y \log x_1} \cdot r_{\log x_1 \log x_2}}{\sqrt{(1 - r^2_{\log y \log x_1})(1 - r^2_{\log x_1 \log x_2})}} = 0{,}663798 .$$

Coeficienţii corelaţiei parţiale arată că legătura dintre $Y$ şi $X_1$ este mai strânsă decât cea dintre $Y$ şi $X_2$.

**Raportul determinaţiei totale** se poate afla prin ridicarea la pătrat a raportului de corelaţie:

$$D_{\log y \log x_1 \log x_2} = \eta^2_{\log y \log x_1 \log x_2} = (0{,}961158)^2 = 0{,}923826 \ (92{,}38\%) .$$

Coeficienţii determinaţiei parţiale se calculează fie prin ridicarea la pătrat a coeficienţilor corelaţiei parţiale:

$$d_{\log y \log x_1 \log x_2} = (0{,}831466\ )^2 = 0{,}691336$$

$$d_{\log y \log x_2 \log x_1} = (0{,}663798\ )^2 = 0{,}440628\ ,$$

fie prin relațiile obținute pe baza divizării determinației multiple totale,

$$D_{yx_1x_2} = d_{yx_2 \cdot x_1} + d_{yx_2 \cdot x_1}\ ,$$

și anume

$$d_{\log y \log x_1 \log x_2} = \frac{b_1 n \sum(\log x_1 \log y) - \sum \log x_1 \sum \log y}{\sqrt{n \sum(\log y)^2 - (\sum \log y)^2}} =$$

$$= \frac{0{,}728269 \cdot 40 \cdot 5{,}16665 - 10{,}7989 \cdot 9{,}82793}{40 \cdot 5{,}49318 - (9{,}82793)^2} = \frac{73{,}2166}{136{,}1396} = 0{,}5945$$

$$d_{\log y \log x_2 \log x_1} = \frac{b_2 n \sum(\log x_2 \log y) - \sum \log x_2 \sum \log y}{\sqrt{n \sum(\log y)^2 - (\sum \log y)^2}} = 0{,}329236\ .$$

Reiese că *92,38%* din variația ponderii producției medii industriale pe județe este datorată influenței simultane a ponderii personalului muncitor și a ponderii fondurilor fixe în industria județelor față de total țară. Influența totală, descompusă pe cei doi factori, arată că ponderea personalului muncitor are o influență mai mare (*59,45%*) în comparație cu ponderea fondurilor fixe în industria județelor față de total țară (*33%*).

### 3.2.3. Modele polinomiale

În cazul unei legături multiple curbilinii de tipul polinomului de gradul doi, ecuația de regresie ia forma:

$$Y_{x_1x_2\ldots} = a + b_1 X_1 + b_1' X_1^2 + b_2 X_2 + b_2' X_2^2 + \ldots + b_p X_p + b_p' X_p^2$$

Pentru o ecuație de regresie de tipul polinomului de gradul doi cu două variabile factoriale:

$$Y_{x_1x_2} = a + b_1 X_1 + b_1' X_1^2 + b_2 X_2 + b_2' X_2^2$$

se obține sistemul de ecuații normale prin rezolvarea căruia se află valoarea parametrilor.

BCU IASI / CENTRAL UNIVERSITY LIBRARY

# Capitolul 4

# TESTAREA IPOTEZELOR MODELULUI CLASIC DE REGRESIE

1. Normalitatea erorilor
2. Homoscedasticitatea
3. Lipsa autocorelării erorilor
4. Lipsa coliniarităţii variabilelor independente
5. Testarea ipotezelor în SPSS

În acest capitol, sunt prezentate principalele probleme legate de testarea ipotezelor unui model de regresie. Sunt prezentate următoarele restricţii de modelare econometrică: normalitatea erorilor, homoscedasticitatea, necorelarea erorilor şi lipsa coliniarităţii variabilelor independente.

Pentru fiecare ipoteză se va utiliza un demers care cuprinde următoarele etape: prezentarea ipotezei, efectele încălcării ipotezei, testarea ipotezei şi eventual procedee de corecţie, în cazul încălcării acesteia.

BCU CLUJ CENTRAL UNIVERSITY LIBRARY

## 4.1. Normalitatea erorilor

Ipoteza de normalitate a erorilor este importantă pentru stabilirea proprietăţilor estimatorilor parametrilor modelului de regresie. Dacă este respectată ipoteza de normalitate a erorilor, adică $\varepsilon_i \sim N(0,\sigma^2)$, estimatorii $\hat{\alpha},\hat{\beta}$ urmează, de asemenea, o lege normală: $\hat{\alpha} \sim N(\beta,\sigma_{\hat{\alpha}}^2)$, $\hat{\beta} \sim N(\beta,\sigma_{\hat{\beta}}^2)$.

### Efectele încălcării ipotezei

Dacă ipoteza de normalitate este încălcată, proprietăţile estimatorilor construiţi pe baza metodei celor mai mici pătrate au doar proprietăţi asimptotice, adică necesită eşantioane sau seturi mari de date.

### Testarea ipotezei de normalitate a erorilor

Pentru testarea normalităţii repartiţiei erorilor se pot utiliza procedee grafice şi procedee numerice. Dintre procedeele grafice amintim: *histograma, curba frecvenţelor, box-plot, Q-Q Plot, P-P Plot.* Ca procedee numerice se utilizează teste: *testul hi-pătrat, testul Kolmogorov* (teste neparametrice), *testul Jarque-Bera.*

### Testul Jarque-Bera

Testul Jarque-Bera este construit pe baza parametrilor formei unei repartiţii: asimetria şi boltirea.

Parametrii formei unei repartiţii sunt: $S = \frac{\mu_3}{\sqrt{\mu_2^3}}$ (S este asimetria – *skewness*, S = 0 pentru o repartiţie normală, $S > 0$ pentru o repartiţie asimetrică la dreapta, respectiv $S < 0$ pentru o repartiţie asimetrică la stânga) şi $K = \frac{\mu_4}{\mu_2^2}$ (K este boltirea – *kurtosis*, $K = 3$ pentru o repartiţie normală, $K < 3$ pentru o repartiţie aplatizată şi $K > 3$ pentru o repartiţie afectată de boltire).

Estimatorii pentru cei doi parametri, în cazul variabilei reziduale, au următoarele relații:

$$\hat{S} = \sqrt{\frac{(\sum_i \frac{\hat{\varepsilon}_i^3}{n-2})^2}{(\sum_i \frac{\hat{\varepsilon}_i^2}{n-2})^3}}, \text{ respectiv, } \hat{K} = \frac{\sum_i \frac{\hat{\varepsilon}_i^4}{n-2}}{(\sum_i \frac{\hat{\varepsilon}_i^2}{n-2})^2}$$

Testul Jarque – Bera are următoarea expresie:

$$JB = \frac{n}{6}\left(\hat{S}^2 + \frac{(\hat{K}-3)^2}{4}\right) \sim \chi^2(2)$$

*Regula de decizie* cu privire la ipoteza de normalitate presupune compararea valorii calculate a testului cu valoarea teoretică. Dacă valoarea calculată este mai mică decât valoarea teoretică, pentru o distribuție hi-pătrat de două grade de libertate și un prag de semnificație specificat, atunci se acceptă ipoteza de normalitate.

Valoarea calculată a testului are la bază estimațiile parametrilor formei repartiției erorilor:

$$s = \sqrt{\frac{(\sum_i \frac{e_i^3}{n-2})^2}{(\sum_i \frac{e_i^2}{n-2})^3}}, \; k = \frac{\sum_i \frac{e_i^4}{n-2}}{(\sum_i \frac{e_i^2}{n-2})^2}, \text{ unde } e_i = y_i - y_{x_i}.$$

Rezultă valoarea calculată a testului: $JB_{calc} = \frac{n}{6}\left(s^2 + \frac{(k-3)^2}{4}\right)$.

## 4.2. Homoscedasticitatea

Homoscedasticitatea presupune o varianță constantă a erorilor $\varepsilon_i$ la nivelul distribuțiilor condiționate, $Y|X = x_i$. Ipoteza de homoscedasticitate cere respectarea relației: $V(\varepsilon_i) = \sigma^2$.

Când ipoteza este încălcată, modelul este numit heteroscedastic.

BCU ASCENTRAUNIVERSITY LIBRARY

### Efectele heteroscedasticităţii

Cel mai important efect al heteroscedasticităţii este pierderea eficienţei estimatorilor parametrilor modelului de regresie. De exemplu, pentru parametrul $\beta$, se poate arăta că varianţa estimatorului suferă modificări în cazul heteroscedasticităţii şi este mai mare decât în cazul în care ipoteza nu este încălcată.

Se poate demonstra că pentru estimatorul parametrului $\beta$ are loc relaţia: $\hat{\beta} = \beta + \sum_i w_i \varepsilon_i$, cu $w_i = \frac{(X_i - \bar{X})}{\sum_i (X_i - \bar{X})^2}$.

Dacă ipoteza de homoscedasticitate este încălcată, adică $V(\varepsilon_i) = \sigma_i^2$, atunci varianţa estimatorului este:

$V(\hat{\beta}) = V(\sum_i w_i \varepsilon_i) = \sum_i w_i^2 V(\varepsilon_i) = \sum_i w_i^2 \sigma_i^2$ şi diferă de varianţa estimatorului, fiind mai mare decât cea în cazul respectării ipotezei de homoscedasticitate.

### Testarea homoscedasticităţii

Pentru testarea homoscedasticităţii se utilizează mai multe teste: testul corelaţiei neparametrice între $\hat{\varepsilon}_i$ şi $X_i$, testul Goldfeld-Quandt, testul Glejser, testul White ş.a.

#### a. Testul corelaţiei neparametrice între $\hat{\varepsilon}_i$ şi $X_i$

Etapele testării sunt următoarele:

– se realizează regresia $Y = \alpha + \beta X + \varepsilon$, fără a ţine seama de ipoteza de homoscedasticitate;

– se estimează erorile $\hat{\varepsilon}_i$ la nivelul eşantionului;

– se calculează rangurile pentru valorile absolute ale erorilor estimate şi pentru valorile variabilei independente;

– se calculează coeficientul de corelaţie al rangurilor Spearman între $|\hat{\varepsilon}_i|$ şi $X_i$;

– se testează coeficientul de corelaţie cu ajutorul testului Student;

– dacă se acceptă ipoteza nulă conform căreia coeficientul de corelaţie nu este semnificativ, se acceptă şi ipoteza de homoscedasticitate, iar în caz contrar, modelul este heteroscedastic.

Relațiile utilizate în acest demers sunt:

- estimatorul coeficientului de corelație: $\hat{\theta} = 1 - \frac{6\sum_i \hat{d}_i^2}{n(n^2-1)}$, unde $\hat{d}_i$ reprezintă estimatorii diferențelor dintre ranguri pentru cele două variabile, iar $n$ este volumul eşantionului;

- testul Student: $t = \frac{\hat{\theta}\sqrt{n-2}}{\sqrt{1-\hat{\theta}^2}} \sim t(n-2)$;

Ipoteza de homoscedasticitate este acceptată când are loc $H_0$: $\theta = 0$, pentru un risc admis.

## Testul Goldfeld-Quandt

Acest test este utilizat în cazul în care volumul eşantionului este suficient de mare. Aplicarea testului presupune parcurgerea următoarelor etape:

– se ordonează crescător seria empirică după variabila $X$;

– se împarte seria în două părți egale, după omiterea unui set de date din centrul seriei. Sensul omiterii acestui set de date este obținerea a două subeşantioane egale ca volum ($n_1 = n_2$) cu date relativ omogene, unul pentru valorile mici ale variabilei şi unul pentru valorile mari ale acesteia;

– se estimează ecuația de regresie pentru fiecare din cele două seturi de date şi se calculează variația reziduală (RSS) pentru fiecare model în parte;

– se realizează un test Fisher care compară cele două variații reziduale, $F = \frac{RSS_2}{RSS_1}$ (raportul are la numărător valoarea cea mai mare pentru variația reziduală, pentru a da o valoare comparabilă cu cea din tabela Fisher);

– valoarea teoretică a testului se citeşte din tabela Fisher pentru $F_{\alpha, n_1-k, n_2-k}$;

– decizia: dacă testul este nesemnificativ statistic, atunci modelul inițial de regresie este homoscedastic, iar în caz contrar avem un model heteroscedastic.

BCU IASI / CENTRAL UNIVERSITY LIBRARY

### Testul Glejser

Acest test are la bază un model de regresie între variabila reziduală estimată, în valoare absolută, şi variabila independentă. Realizarea testului presupune parcurgerea următorului demers:

– se construieşte modelul de regresie $y_i = \alpha + \beta x_i + \varepsilon_i$;

– se determină erorile estimate $\hat{\varepsilon}_i$;

– se construieşte un model de regresie între $|\hat{\varepsilon}_i|$ şi $X_i$. De exemplu, se realizează regresia liniară $|\hat{\varepsilon}_i| = \alpha_0 + \alpha_1 x_i + u_i$;

– se testează modelul de regresie $|\hat{\varepsilon}_i| = \alpha_0 + \alpha_1 x_i + u_i$. Dacă parametrul $\alpha_1$ nu diferă semnificativ de zero, atunci modelul iniţial este homoscedastic.

### Corectarea modelului heteroscedastic

Dacă modelul de regresie este heteroscedastic, are loc relaţia: $V(\varepsilon_i) = \sigma_i^2$.

Corectarea heteroscedasticităţii se poate realiza ţinând cont de informaţia disponibilă asupra parametrilor $\sigma_i^2$. Există două situaţii: dacă sunt cunoscuţi parametrii $\sigma_i^2$ sau dacă aceştia sunt necunoscuţi.

### a. Cazul $\sigma_i^2$ cunoscuţi.

### Metoda celor mai mici pătrate generalizată

Se consideră modelul de regresie $y_i = \alpha + \beta x_i + \varepsilon_i$. Dacă se cunosc $\sigma_i^2$, modelul poate fi rescris în următoarea formă:

$$\frac{y_i}{\sigma_i} = \frac{\alpha}{\sigma_i} + \beta \frac{x_i}{\sigma_i} + \frac{\varepsilon_i}{\sigma_i} \text{ sau } y_i^* = \alpha^* \frac{1}{\sigma_i} + \beta^* x_i^* + \varepsilon_i^*, \text{ în care:}$$

$$y_i^* = \frac{y_i}{\sigma_i}, \quad \alpha^* = \alpha, \quad \beta^* = \beta, \quad x_i^* = \frac{x_i}{\sigma_i}, \quad \varepsilon_i^* = \frac{\varepsilon_i}{\sigma_i}.$$

Noul model obţinut (modelul „*“) sau modelul modificat este un model homoscedastic, deoarece $V(\varepsilon_i^*) = 1$.

Într-adevăr, $V(\varepsilon_i^*) = V(\frac{\varepsilon_i}{\sigma_i}) = \frac{1}{\sigma_i^2} V(\varepsilon_i) = 1$.

*Observaţie*

Modelul transformat se obţine prin ponderarea modelului iniţial cu o greutate specifică sau cu o pondere de forma: $w_i = \frac{1}{\sigma_i}$.

### b. Cazul $\sigma_i^2$ necunoscuţi

În acest caz, asupra parametrilor $\sigma_i^2$ se emit diferite ipoteze funcţionale, cum ar fi, de exemplu, $\sigma_i^2 = \sigma^2 x_i^2$ sau $\sigma_i^2 = \sigma^2 x_i$ etc. Astfel de relaţii pot fi detectate cu ajutorul testului Glejser, prin construirea de modele de regresie între variabila reziduală şi cea independentă.

Dacă se foloseşte ipoteza $\sigma_i^2 = \sigma^2 x_i^2$, de exemplu, atunci modelul corectat are forma: $\frac{y_i}{x_i} = \frac{\alpha}{x_i} + \beta + \frac{\varepsilon_i}{x_i}$ sau $y_i^* = \alpha^* \frac{1}{x_i} + \beta^* + \varepsilon_i^*$, în care $V(\varepsilon_i^*) = \sigma^2$.

Ponderarea modelului în acest caz se realizează cu ajutorul coeficienţilor: $w_i = \frac{1}{x_i}$.

## 4.3. Lipsa autocorelării erorilor

Ipoteza de necorelare a erorilor se referă la lipsa unei corelaţii între erorile $\varepsilon_i$ la nivelul distribuţiilor condiţionate, de forma $Y \mid X = x_i$. Cu alte cuvinte, ipoteza presupune că eroarea asociată unei valori a variabilei dependente nu este influenţată de eroarea asociată altei valori a variabilei dependente.

Dacă nu se respectă această ipoteză, suntem în situaţia existenţei fenomenului de *autocorelare* a erorilor sau a *corelaţiei seriale*, adică $cov(\varepsilon_i, \varepsilon_j) \neq 0$ sau $M(\varepsilon_i, \varepsilon_j) \neq 0$.

Autocorelarea erorilor poate să apară din diferite cauze: inerţia fenomenelor în timp şi decalajul, în cazul seriilor de timp, lipsa unei specificaţii adecvate a formei modelului de regresie, lipsa din model a uneia

sau a mai multor variabile semnificative, sistematizarea şi pregătirea datelor pentru prelucrare.

În cazul existenţei fenomenului de autocorelare, considerăm că între erori există o relaţie dată de modelul:

$$\varepsilon_i = \rho\varepsilon_{i-1} + u_i,$$

unde $u_i$ reprezintă o variabilă aleatoare, numită „zgomot alb" sau variabilă pur aleatoare, care respectă ipotezele modelului clasic de regresie, iar $\rho$ este coeficientul de corelaţie sau de autocorelaţie între variabilele $\varepsilon_i$ şi $\varepsilon_{i-1}$.

Pentru variabila $u_i$ au loc relaţiile:

- $M(u_i)=0$;
- $V(u_i)=\sigma_u^2$, ipoteza de homoscedasticitate;
- $u_i \sim N(0,\sigma_u^2)$, ipoteza de normalitate;
- $cov(u_i,u_j)=0$, ipoteza de necorelare a erorilor.

*Coeficientul de autocorelaţie de ordinul I*

Coeficientul de autocorelaţie între $\varepsilon_i$ şi $\varepsilon_{i-1}$ este definit prin relaţia:

$$\rho = \frac{cov(\varepsilon_i,\varepsilon_{i-1})}{\sigma_i\sigma_{i-1}} = \frac{cov(\varepsilon_i,\varepsilon_{i-1})}{\sigma^2} = \frac{\sum_i \varepsilon_i \cdot \varepsilon_{i-1}}{\sum_i \varepsilon_i^2}.$$

În aceste condiţii, $\sigma^2 = V(\varepsilon_i) = V(\rho\varepsilon_{i-1} + u_i) = \rho^2\sigma^2 + \sigma_u^2$, de unde rezultă $\sigma^2 = \frac{\sigma_u^2}{1-\rho^2}$.

*Observaţie*

Relaţia de mai sus ne indică maniera în care fenomenul de autocorelare a erorilor afectează dispersia acestora: lipsa autocorelării ($\rho=0$) presupune o dispersie specifică unei variabile reziduale pur aleatoare ($\sigma^2 = \sigma_u^2$), respectiv, o corelaţie perfectă între erori ($\rho=1$) determină o dispersie infinită a acestora.

BCU TRANSILVANIA CENTRALA UNIVERSITARA LIBRARY

*Coeficientul de autocorelaţie de ordinul k – funcţia de autocorelaţie*

Coeficientul de autocorelare se poate determina şi între două valori între care există un decalaj cu ordin mai mare decât unu. Pentru astfel de situaţii, se defineşte funcţia de autocorelaţie, potrivit relaţiei:

$$f(k)=\frac{cov(\varepsilon_i,\varepsilon_{i-k})}{\sigma_i\sigma_{i-k}}=\frac{cov(\varepsilon_i,\varepsilon_{i-k})}{\sigma^2}.$$

## Efectele încălcării ipotezei

În cazul existenţei autocorelării erorilor, este afectată calitatea estimatorilor parametrilor modelului de regresie, în special, estimatorii îşi pierd eficienţa.

Prin încălcarea ipotezei de necorelare a erorilor, se poate demonstra că pentru parametrul $\alpha$, prin aplicarea metodei celor mai mici pătrate, se obţine un estimator nedeplasat, dar neeficient.

Se poate arăta că pentru parametrul $\alpha$ se obţine un alt estimator, $\hat{\alpha}^*$, care are varianţă minimă.

Considerăm modelul liniar de regresie: $y_i=\alpha+\beta x_i+\varepsilon_i$, pentru care are loc relaţia $\varepsilon_i=\rho\varepsilon_{i-1}+u_i$. Modelul pentru un decalaj de ordinul unu este dat de relaţia: $y_{i-1}=\alpha+\beta x_{i-1}+\varepsilon_{i-1}$. Dacă acest ultim model se înmulţeşte cu $\rho$ şi se scade din modelul iniţial, rezultă modelul:

$$y_i-\rho\cdot y_{i-1}=\alpha(1-\rho)+\beta(x_i-\rho\cdot x_{i-1})+u_i$$

Scris altfel, modelul de mai sus este de forma:

$y_i^*=\alpha^*+\beta^* x_i^*+u_i$, numit şi model de *quasi-diferenţă,* unde $\alpha^*=\alpha(1-\rho)$, $\beta^*=\beta$, $x_i^*=x_i-\rho\cdot x_{i-1}$, $u_i=\varepsilon_i-\rho\varepsilon_{i-1}$

Acest model respectă ipotezele modelului clasic de regresie, iar prin aplicarea metodei celor mai mici pătrate ne oferă un alt estimator pentru parametrul $\alpha$, estimator care este nedeplasat şi eficient.

## Testarea autocorelării erorilor

În literatura de specialitate, se întâlnesc mai multe teste pentru această ipoteză, iar dintre acestea vom prezenta două: *Runs test* şi *Durbin Watson test.*

### a. Runs test

Acest test are la bază ideea că valorile variabilei reziduale se constituie în secvențe sau seturi de valori pozitive sau negative numite *runs*, care se succed într-o anumită ordine sau aleator. De exemplu, o variabilă eroare cu 10 valori poate fi dată de următoarea succesiune de 3 runs:

(- - -) (+ + +) (- - - -).

Ipoteza de bază a acestui test este aceea că în cazul lipsei autocorelării erorilor succesiunea de runs este aleatoare sau numărul acestora este distribuit normal.

*Notații*

Pentru un eșantion de volum $n$, se fac următoarele notații:

- $\hat{\varepsilon}$ este variabila care estimează erorile la nivelul eșantionului;
- $n_1$ este numărul de valori pozitive ale variabilei $\hat{\varepsilon}$;
- $n_2$ este numărul de valori negative ale variabilei $\hat{\varepsilon}$, $n_1 + n_2 = n$;
- $k$ este numărul de runs, variabilă aleatoare care are următorii parametri:

$$M(k) = 2\frac{n_1 n_2}{n_1 + n_2} + 1,$$

$$V(k) = \sigma_k^2 = 2n_1 n_2 \frac{2n_1 n_2 - n_1 - n_2}{(n_1 + n_2)^2 (n_1 + n_2 - 1)}.$$

**Etapele testării**

1. Formularea ipotezelor:

$H_0$: $k$ este distribuit normal (nu există autocorelare a erorilor);

$H_1$: $k$ nu este distribuit normal (ipoteza este încălcată);

2. se utilizează o statistică Student: $t = \dfrac{k - M(k)}{\hat{\sigma}_k}$;

3. decizia de a accepta ipoteza nulă se ia comparând valoarea calculată a testului la nivelul unui eșantion, cu valoarea teoretică, corespunzătoare unui risc asumat.

O variantă pentru a lua decizia de a admite sau de a respinge ipoteza de necorelare a erorilor, pentru acest test, este să verificăm relația:

$M(k) - 1{,}96\hat{\sigma}_k \le k \le M(k) + 1{,}96\hat{\sigma}_k$, pentru un prag de semnificație de 0,05, de exemplu.

*Observaţie*

În SPSS, pentru determinarea numărului de runs se utilizează semnul diferenţei dintre fiecare valoare a variabilei eroare estimată şi media acestora sau a unei alte mărimi medii, cum ar fi mediana sau modul.

### b. Durbin Watson test

Acest test are la bază ideea testării semnificaţiei coeficientului de corelaţie de ordinul I a erorilor, $\rho$.

Etapele testului:

**1.** Formularea ipotezelor

$H_0$: $\rho = 0$ (nu există autocorelare a erorilor);

$H_1$: $\rho \neq 0$ (ipoteza este încălcată);

**2.** Statistica test utilizată este: $DW = d = \dfrac{\sum_i (\hat{\varepsilon}_i - \hat{\varepsilon}_{i-1})^2}{\sum_i \hat{\varepsilon}_i^2}$.

În condiţiile în care $\varepsilon_i = \rho\varepsilon_{i-1} + u_i$, statistica $DW$ se mai poate scrie astfel:

$$d = \frac{\sum_i \hat{\varepsilon}_i^2 - 2\sum_i \hat{\varepsilon}_i\hat{\varepsilon}_{i-1} + \sum_i \hat{\varepsilon}_{i-1}^2}{\sum_i \hat{\varepsilon}_i^2} \cong 2\frac{\sum_i \hat{\varepsilon}_i^2 - \sum_i \hat{\varepsilon}_i\hat{\varepsilon}_{i-1}}{\sum_i \hat{\varepsilon}_i^2} = 2\left(1 - \frac{\sum_i \hat{\varepsilon}_i\hat{\varepsilon}_{i-1}}{\sum_i \hat{\varepsilon}_i^2}\right) = 2(1-\hat{\rho})$$

unde $\hat{\rho} = \dfrac{\sum_i \hat{\varepsilon}_i\hat{\varepsilon}_{i-1}}{\sum_i \hat{\varepsilon}_i^2}$.

În concluzie, statistica $DW$ se poate aproxima prin relaţia: $d = 2(1-\hat{\rho})$.

Deoarece $-1 \le \hat{\rho} \le 1$, valorile statisticii $DW$ sunt date de intervalul: $0 \le d \le 4$. Astfel, au loc rezultatele:

$\hat{\rho} = -1 \Rightarrow d = 4$, există autocorelare negativă maximă a erorilor;

$\hat{\rho} = 1 \Rightarrow d = 0$, există autocorelare pozitivă maximă a erorilor;

$\hat{\rho} = 0 \Rightarrow d = 2$, nu există autocorelare.

**3.** Decizia se ia pe baza valorilor critice ale statisticii $DW$, calculate şi tabelate în funcţie de pragul de semnificaţie şi de volumul eşantionului. În tabele se determină două valori critice, notate cu $d_L$ (limita inferioară) şi $d_U$

(limita superioară), pentru diverse valori ale pragului de semnificație, ale numărului de parametri din model și ale volumului eșantionului. În funcție de aceste valori critice se determină următoarele intervale, care permit luarea deciziei de respingere sau de acceptare a ipotezei nule:

Decizia se ia în funcție de poziția valorii calculate a testului:

*(0; $d_L$)* este o regiune de respingere, erorile înregistrează o autocorelare pozitivă;

*($d_L$; $d_U$)* și *(4-$d_u$; 4-$d_L$)* sunt regiune de nedeterminare, dacă valoarea statisticii *DW* cade în această regiune, nu se poate decide asupra existenței autocorelării erorilor;

*($d_u$ ; 4- $d_u$)* este o regiune de acceptare a ipotezei nule, erorile nu sunt autocorelate;

*(4-$d_L$ ; 4)* este o regiune de respingere, erorile înregistrează o autocorelare negativă.

*Observație*

Testul Durbin Watson nu realizează decât o testare asupra existenței unei autocorelări de ordinul întâi între termenii variabilei eroare.

## Corectarea autocorelării erorilor

Considerăm modelul liniar de regresie $y_i = \alpha + \beta x_i + \varepsilon_i$.

Există două situații posibile pentru corectarea autocorelării erorilor: când se cunoaște coeficientul de autocorelație dintre erori și când acesta nu se cunoaște.

### a. $\rho$ este cunoscut

În acest caz, estimarea parametrilor modelului se realizează cu ajutorul modelului de regresie modificat, adică a modelului de *quasi-diferență:* $y_i - \rho \cdot y_{i-1} = \alpha(1-\rho) + \beta(x_i - \rho \cdot x_{i-1}) + u_i$ sau

$y_i^* = \alpha^* + \beta^* x_i^* + u_i$, unde

$Y_i^* = Y_i - \rho Y_{i-1}$, $\quad \alpha^* = \alpha(1-\rho)$

$X_i^* = X_i - \rho X_{i-1}$, $\quad \beta^* = \beta$

BCU "LUCIAN BLAGA" UNIVERSITY LIBRARY

$u_i = \varepsilon_i - \rho\varepsilon_{i-1}$

Pentru modelul (*) există doi estimatori nedeplasați, convergenți și eficienți, $\hat{\alpha}^*, \hat{\beta}^*$, care se determină cu ajutorul metodei celor mai mici pătrate. În aceste condiții, estimatorii pentru parametrii modelului inițial sunt:

$$\hat{\alpha} = \frac{\hat{\alpha}^*}{1-\rho}, \quad \hat{\beta} = \hat{\beta}^*.$$

### b. $\rho$ este necunoscut

În acest caz, există mai multe metode de estimare a parametrilor modelului inițial care au la bază estimarea coeficientului de autocorelație dintre erori. O metodă larg utilizată este procedeul iterativ Cochrane-Orcutt.

### Etapele procedeului Cochrane-Orcutt

Ca proces iterativ, procedeul Cochrane-Orcutt presupune realizarea unor iterații, iar fiecare iterație presupune parcurgerea unui număr de pași. Procesul se oprește după un număr de iterații prin respectarea unei condiții.

O iterație presupune următorii pași:

*Pas 1.* Realizarea regresiei $y_i = \alpha + \beta x_i + \varepsilon_i$, fără a ține cont de ipoteza de necorelare a erorilor și estimarea erorilor $\hat{\varepsilon}_i$.

*Pas 2.* Realizarea regresiei $\hat{\varepsilon}_i = \hat{\rho}\hat{\varepsilon}_{i-1} + v_i$ și calcularea unei valori estimate pentru parametrul $\rho$.

*Pas 3.* Realizarea unei regresii pe baza modelului de *quasi-diferență* $y_i^* = \alpha^* + \beta^* x_i^* + u_i$, utilizând estimația coeficientului de autocorelație de la pasul 2.

*Pas 4.* Procedeul se încheie dacă se respectă o condiție de oprire. În caz contrar, se reia pasul 1, într-o nouă iterație, cu rezultatele de la pasul 3.

Procedeul se oprește atunci când între două valori estimate ale coeficientului de autocorelație din două iterații succesive verifică relația: $\rho^{(k)} - \rho^{(k-1)} = 0{,}0025$.

BCU STEFAN CEL MARE UNIVERSITY LIBRARY

## 4.4. Lipsa coliniarităţii variabilelor independente

Ipoteza de necoliniaritate presupune că între variabilele independente ale unui model de regresie nu există o legătură de tip liniar. Principalele probleme legate de această ipoteză sunt: identificarea gradului de coliniaritate, stabilirea cauzelor încălcării ipotezei, stabilirea efectelor coliniarităţii, testarea ipotezei de coliniaritate şi corectarea modelului în cazul existenţei acesteia.

### Grade de coliniaritate

Considerăm un model de regresie care are $p$ variabile independente: $X_1, X_2, ..., X_p$. Între variabilele independente există o coliniaritate *perfectă* dacă există $p$ constante $\lambda_1, \lambda_2, ..., \lambda_p$, nu toate nule, astfel încât să aibă loc relaţia:

$$\lambda_1 X_1 + \lambda_2 X_2 + ... + \lambda_p X_p = 0$$

Coliniaritate dintre variabile se numeşte *neperfectă*, dacă are loc relaţia:

$\lambda_1 X_1 + \lambda_2 X_2 + ... + \lambda_p X_p + u = 0$, unde $u$ este o variabilă aleatoare care respectă ipotezele modelului clasic de regresie.

### Cauzele coliniarităţii

Coliniaritatea poate apare din diferite surse: natura fenomenului şi variabilele alese pentru a realiza modelarea (de exemplu, într-un model asupra consumului de bunuri de strictă necesitate, variabile independente ca venitul şi mărimea familiei pot fi corelate liniar), tipul de model de regresie utilizat etc. Fenomenul de coliniaritate apare la nivelul eşantionului de date disponibile, nu la nivelul populaţiei totale.

### Efectele încălcării ipotezei

Efectul coliniarităţii se evidenţiază în varianţa estimatorilor parametrilor modelului de regresie. În cazul unei coliniarităţi perfecte, parametrii asociaţi variabilelor independente între care există o legătură de tip liniar perfectă nu pot fi determinaţi, iar varianţa acestora este practic infinită. În cazul unei coliniarităţi imperfecte, coeficienţii de regresie ai variabilelor independente coliniare se estimează cu o eroare foarte mare.

## Testarea coliniarităţii

În practică, identificarea coliniarităţii variabilelor independente se poate realiza prin diferite metode:

**a.** Testarea coeficienţilor de regresie în cazul unui model cu un coeficient de determinaţie ridicat (de obicei peste 0.8). Dacă coeficienţii de regresie sunt nesemnificativ diferiţi de zero, atunci ipoteza de necoliniaritate este încălcată.

**b.** Testarea coeficienţilor de corelaţie bivariaţi pentru variabilele independente din modelul de regresie. Dacă aceşti coeficienţi au valori ridicate (de regulă, peste 0.8), atunci există posibilitatea coliniarităţii între variabilele independente.

**c.** Estimarea şi testarea parametrilor modelelor de regresie auxiliară dintre variabilele independente. Ipoteza de necoliniaritate este încălcată dacă aceşti coeficienţi de regresie sunt semnificativ diferiţi de zero.

În software-urile de statistică, cum este şi cazul pachetului SPSS, se utilizează o metodă de detectare a coliniarităţii pe baza a doi indicatori: *Tolerance* şi *VIF (Variance Inflation Factor)*. Ideea de bază a acestei metode se bazează pe câteva proprietăţi specifice unui model liniar multiplu.

Considerăm modelul de regresie liniară multiplă :

$$y_i = \alpha + \beta_1 x_{1i} + \beta_2 x_{2i} + \varepsilon_i .$$

Pentru indicatorii de mai sus, au loc relaţiile:

$$V(\hat{\beta}_1) = \frac{\sigma^2}{\sum_i (x_{1i} - \bar{x}_1)^2 (1 - r_{12}^2)}, \quad V(\hat{\beta}_2) = \frac{\sigma^2}{\sum_i (x_{2i} - \bar{x}_2)^2 (1 - r_{12}^2)},$$

în care:

$$r_{12}^2 = \frac{\left[\sum_i (x_{1i} - \bar{x}_1)(x_{2i} - \bar{x}_2)\right]^2}{\left[\sum_i (x_{1i} - \bar{x}_1)^2\right]\left[\sum_i (x_{2i} - \bar{x}_2)^2\right]}$$

şi reprezintă coeficientul de determinaţie din modelul de regresie auxiliar bazat pe cele două variabile independente.

Dacă generalizăm, putem scrie o relaţie pentru dispersia estimatorilor parametrilor modelului de regresie:

$V(\hat{\beta}_j) = \frac{\sigma^2}{\sum_i (x_{ji} - \bar{x}_j)^2 (1 - r_j^2)}$, unde $r_j^2$ este raportul de determinație din modelul de regresie auxiliar, construit pe baza variabilelor independente, în care variabila *j* este considerată variabila dependentă, iar celelalte variabile factoriale sunt considerate variabile independente.

Indicatorul *VIF* se definește prin relația:

$$VIF_j = \frac{1}{(1 - r_j^2)}.$$

Acesta indică modul în care varianța estimatorului unui coeficient de regresie este influențată de prezența coliniarității la nivelul variabilelor independente. Evident, lipsa coliniarității dă o valoare *VIF = 1,* iar existența coliniarității determină o valoare mare a indicatorului, condiția limită fiind în cazul unei coliniarități perfecte ($r_j^2 = 1, \Rightarrow VIF \to \infty$).

În practică, se consideră că o valoare *VIF>10* indică prezența coliniarității.

Indicatorul *Tolerance* se determină ca inversul valorii indicatorului *VIF.* Astfel, are loc relația:

$$TOL_j = \frac{1}{VIF_j} = (1 - r_j^2).$$

Dacă *TOL = 1,* nu există coliniaritate, iar dacă *TOL = 0* suntem în situația extremă, de coliniaritate perfectă.

## Corectarea coliniarității

Pentru corectarea coliniarității se iau în calcul următoarele probleme: dimensiunile coliniarității, informațiile suplimentare cunoscute *a priori*, numărul de variabile independente și importanța lor etc. În acord cu aceste condiții, se pot aplica mai multe metode de corecție:

– Eliminarea variabilei care induce fenomenul de coliniaritate. Această metodă necesită însă destulă precauție, pentru a nu comite o eroare de specificare a modelului.

– Combinarea datelor din anchete cu date din serii de timp (*pooling the data*). Prin această metodă se poate reduce influența unei variabile independente asupra celorlalte, estimând coeficientul de regresie al variabilei respective pe baza unui alt set de date disponibile.

– Transformarea variabilelor modelului prin diferite procedee, cum ar fi aplicarea operatorilor *lag* (decalaj) sau diferență.

– Mărirea numărului eşantionului sau includerea de noi date în analiză. Deoarece coliniaritatea este dată în special de construcția eşantionului, includerea de noi date poate reduce sensibil fenomenul.

## 4.5. Testarea ipotezelor în SPSS

Pentru exemplificarea demersului testării ipotezelor modelului clasic de regresie, se consideră un exemplu de model de regresie liniară multiplă, pe baza datelor disponibile din statistica oficială (Anuarul Statistic al României, 2001).

### Variabile

Variabilele din exemplu sunt prezentate în lista din figura 1. Variabila dependentă este *valoarea serviciilor din agricultură.*

bd_initiala - SPSS Data Editor

File Edit View Data Transform Analyze Graphs Utilities Window Help

| | Name | Type | Width | Decimals | Label | Values |
|---|---|---|---|---|---|---|
| 1 | judet | String | 20 | 0 | judetul | None |
| 2 | regiunea | Numeric | 8 | 0 | regiunea | {1, nord-est} |
| 3 | pop_ocup | Numeric | 8 | 2 | populatia ocupata in agricultura(mii persoane) | None |
| 4 | nrm_sal | Numeric | 8 | 0 | nr mediu al salariatilor in agricultura(mii persoane) | None |
| 5 | sal_m | Numeric | 8 | 0 | salariul mediu lunar in agricultura(lei/salariat) | None |
| 6 | suprf_c | Numeric | 8 | 0 | suprafata cultivata(ha) | None |
| 7 | prod_agr | Numeric | 8 | 0 | productia agricola vegetala(milioane lei) | None |
| 8 | serv_agr | Numeric | 8 | 0 | valoarea serviciilor agricole(mil lei) | None |
| 9 | suprf_ag | Numeric | 8 | 0 | suprafata agricola(ha) | None |
| 10 | | | | | | |
| 11 | | | | | | |

**Figura nr. 1.** *Variabile utilizate în exemplu*

În urma analizei specifice modelelor liniare multiple, după demersul prezentat la capitolul 3, se obține un model liniar multiplu cu două variabile independente, așa cum o arată figura 2.

BCU "M. EMINESCU" IAŞI

Variables Entered/Removed[b]

| Model | Variables Entered | Variables Removed | Method |
|---|---|---|---|
| 1 | suprafata cultivata(ha), nr mediu al salariatilor in agricultura (mii persoane)[a] | . | Enter |

a. All requested variables entered.

b. Dependent Variable: valoarea serviciilor agricole(mil lei)

**Figura nr. 2.** *Variabilele din modelul supus testării*

Prin aplicarea demersului de estimare a parametrilor modelului liniar multiplu, cu variabilele prezentate mai sus, se obţin rezultatele din figura 3, care oferă estimaţiile parametrilor modelului.

Coefficients[a]

| Model | | Unstandardized Coefficients | | Standardized Coefficients | t | Sig. |
|---|---|---|---|---|---|---|
| | | B | Std. Error | Beta | | |
| 1 | (Constant) | -13515,9 | 6798,191 | | -1,988 | ,054 |
| | nr mediu al salariatilor in agricultura(mii persoane) | 8,080 | 2,132 | ,483 | 3,790 | ,001 |
| | suprafata cultivata(ha) | ,163 | ,047 | ,447 | 3,502 | ,001 |

a. Dependent Variable: valoarea serviciilor agricole(mil lei)

**Figura nr. 3.** *Estimaţiile coeficienţilor de regresie*

Modelul estimat este:

$$Y = -13515{,}9 + 8{,}080 \cdot nrm\_sal + 0{,}163 \cdot suprf\_c$$

Acest model va fi utilizat în continuare pentru testarea ipotezelor.

*Observaţie*

Ipotezele cu privire la variabila reziduală se vor verifica cu ajutorul estimaţiilor calculate în urma estimării parametrilor modelului de regresie. În SPSS, există posibilitatea salvării acestei variabile, aşa cum arată figura 4.

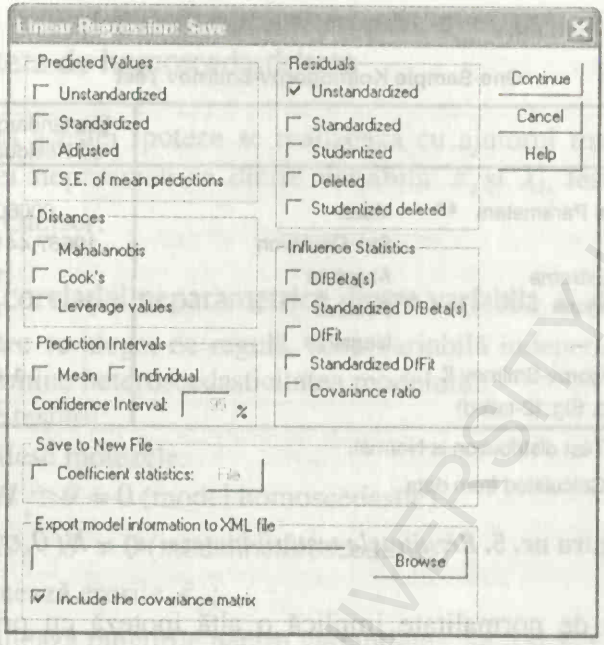



**Figura nr. 4.** *Opțiunea de a salva variabila reziduală estimată*

## a. Ipoteza de normalitate a erorilor

Pentru acest exemplu, vom testa ipoteza de normalitate a erorilor cu ajutorul testului neparametric Kolmogorov-Smirnov (testul K-S).

**Testul K – S** verifică normalitatea repartiției comparând frecvențele cumulate cu frecvențele teoretice cumulate calculate pe baza tabelului Gauss-Laplace.

În SPSS, verificarea normalității cu ajutorul testului K-S presupune urmatorul demers:

meniul ***Analyze*** → comanda ***Nonparametric Tests*** → opțiunea ***One-Sample Kolmogorov – Smirnov Test.*** În fereastra *Test Variable List* introducem variabila reziduală „res_1".

Decizia de a accepta sau respinge ipoteza nulă: $H_0 : \varepsilon_i \sim N(0, \sigma^2)$ se ia pe baza rezultatelor testului, prezentate în figura 5.

*Regula de decizie:*

*sig = 0,16 > 0,05*, prin urmare se acceptă ipoteza $H_0$ cu o probabilitate de *0,95*. În concluzie, distribuția erorilor urmează o lege normală. Astfel, ipoteza de normalitate a erorilor este verificată.

**One-Sample Kolmogorov-Smirnov Test**

| | | Unstandardized Residual |
|---|---|---|
| N | | 40 |
| Normal Parameters [a,b] | Mean | ,0000000 |
| | Std. Deviation | 19667,44153 |
| Most Extreme Differences | Absolute | ,176 |
| | Positive | ,176 |
| | Negative | -,145 |
| Kolmogorov-Smirnov Z | | 1,116 |
| Asymp. Sig. (2-tailed) | | ,166 |

a. Test distribution is Normal.

b. Calculated from data.

**Figura nr. 5.** *Rezultatele testării ipotezei* $\varepsilon_i \sim N(0,\sigma^2)$

Ipoteza de normalitate implică o altă ipoteză cu privire la media erorilor, şi anume $M(\varepsilon_i)=0$. Această ipoteză semnifică faptul că influenţa tuturor factorilor neincluşi în model nu trebuie să afecteze sistematic media variabilei dependente.

În SPSS, pentru testarea acestei ipoteze, se urmează demersul:

meniul ***Analyze*** → comanda ***Compare Means*** → opţiunea ***One-Sample T Test.***

În caseta *Test Variable(s)* se introduce variabila reziduală nestandardizată estimată („res_1") obţinută în urma aplicării demersului regresiei. Această valoare se compară cu 0, deci în caseta *Test Value* se va introduce valoarea 0.

Conform figurii 6, se obţine o valoare a semnificaţiei testului *sig t*= 1,00, deci *sig t > 0,05*, ceea ce înseamnă că se va accepta ipoteza $H_0$, conform căreia media erorilor este zero.

**One-Sample Test**

| | Test Value = 0 | | | | | |
|---|---|---|---|---|---|---|
| | | | | | 95% Confidence Interval of the Difference | |
| | t | df | Sig. (2-tailed) | Mean Difference | Lower | Upper |
| Unstandardized Residual | ,000 | 39 | 1,000 | ,0000000 | -6289,95 | 6289,953 |

**Figura nr. 6.** *Rezultatele testării ipotezei* $M(\varepsilon_i)=0$

BCU IAŞI CENTRAL UNIVERSITY LIBRARY

### b. Ipoteza de homoscedasticitate

Testarea acestei ipoteze se realizează cu ajutorul mai multor teste: testul corelației neparametrice dintre variabila $\hat{\varepsilon}_i$ și $X_i$, testul Goldfeld – Quandt, testul Glejser.

#### Testul corelației neparametrice dintre variabila $\hat{\varepsilon}_i$ și $X_i$

În testare se alege, de regulă, acea variabilă independentă care este posibil să determine heteroscedasticitatea modelului.

Etapele testării:

- se stabilesc ipotezele:

$H_0: \theta = 0$ (model homoscedastic)

$H_1: \theta \neq 0$ (model heteroscedastic)

- se estimează erorile $\hat{\varepsilon}_i$;
- se calculează rangurile pentru variabilele $|\hat{\varepsilon}_i|$ și $X_i$;
- se determină coeficientul de corelație neparametrică Spearman pe baza estimatorului:

$$\hat{\theta} = 1 - \frac{6\sum_{i=1}^{n} \hat{d}_i^2}{n(n^2-1)} \quad ; \quad d_i = R_{|\hat{\varepsilon}_i|} - R_{X_i}$$

- se realizează testarea cu ajutorul testului $t = \frac{\hat{\theta}}{\sqrt{1-\hat{\theta}^2}}\sqrt{n-2} \sim t(n\text{-}2)$
- se calculează $t_0$ și se compară cu valoarea teoretică $t_{\alpha/2}$
- regula de decizie
  - dacă $sig\ t < \alpha$ se respinge $H_0$, adică modelul este heteroscedastic;
  - dacă $sig\ t > \alpha$ se accepta $H_0$, adică modelul este homoscedastic.

În SPSS, construim mai întâi variabila $|\hat{\varepsilon}_i|$.

Ca demers: folosim din bara de meniu ***Transform*** → comanda ***Compute*** (funcția ABS) și noua variabila „res_2", care va fi egală cu |res_1|. Apoi urmăm demersul ***Analyze*** → comanda ***Correlate*** → opțiunea ***Bivariate***. În caseta *Variables* se selectează „suprf_c" (variabila independentă) și noua variabilă „res_2". Se obține următorul output:

Correlations

| | | | RES_2 | suprafata cultivata(ha) |
|---|---|---|---|---|
| Spearman's rho | RES_2 | Correlation Coefficient | 1,000 | ,201 |
| | | Sig. (2-tailed) | . | ,010 |
| | | N | 40 | 40 |
| | suprafata cultivata(ha) | Correlation Coefficient | ,201 | 1,000 |
| | | Sig. (2-tailed) | ,10 | . |
| | | N | 40 | 40 |

**Figura nr. 7.** *Rezultatele testării ipotezei* $H_0 : \theta = 0$

Deoarece *Sig t = 0,101 > 0,05* se acceptă ipoteza $H_0$ cu o probabilitate de 95%, adică modelul este homoscedastic.

## Testul Goldfeld – Quandt

*Etapele testului:*

- se stabilesc ipotezele

  $H_0 : \theta = 0$ (model homoscedastic)

  $H_1 : \theta \neq 0$ (model heteroscedastic)

- se ordonează seria după variabila independentă *X;*
- se împarte seria de date în două părţi după ce au fost excluse un număr de înregistrări din centrul seriei;
- se realizează câte o regresie pentru fiecare din cele două subeşantioane obţinute şi se calculează valoarea estimată a variaţie reziduale (*RSS*) pentru fiecare regresie: $RSS_1$ şi $RSS_2$ ;
- pentru testare se utilizează testul Fischer: $F = \frac{RSS_1}{RSS_2}$ (dacă *F* este semnificativ, există diferenţe semnificative la nivelul celor două seturi de date, deci modelul este heteroscedastic, iar dacă *F* este nesemnificativ modelul este homoscedastic);

BCU "CAROL I" UNIVERSITY LIBRARY

- regula de decizie:

- dacă $sig\ F < \alpha$, se respinge $H_0$, adică modelul este heteroscedastic şi va trebui corectat;

- dacă $sig\ F > \alpha$, se accepta $H_0$, adică modelul este homoscedastic.

În SPSS, ordonăm seria după variabila independentă „suprf_c" (suprafaţa cultivată) şi se vor construi două subeşantioane de volume egale (20 de înregistrări): unul corespunzător valorilor mici ale variabilei independente şi altul corespunzător valorilor mari.

Cele două subeşantioane se construiesc parcurgând următorul demers:

- Meniul ***Transform*** → comanda ***Recode*** → opţiunea ***Recode Into Different Variables.*** Selectând variabila de interes „suprf_c", aceasta va primi o nouă denumire, „suprc2" şi o etichetă nouă: mărimea suprafeţei cultivate definită ca variabila dummy. Valorile mai mici decât 177 000 vor primi codul 0, iar celelalte vor primi codul 1.

- Se utilizează comanda ***Split File*** din meniul ***Data***, pentru a delimita cele două seturi distincte de date. Toate comenzile ulterioare se vor realiza pentru cele două seturi distincte de date.

- Se realizează regresia între variabilele considerate şi se obţin rezultatele din figura 8.

ANOVA[b]

| SUPRC2 | Model | | Sum of Squares | df | Mean Square | F | Sig. |
|---|---|---|---|---|---|---|---|
| ,00 | 1 | Regression | 860121056,041 | 1 | 860121056,041 | 12,247 | ,003[a] |
| | | Residual | 1264204006,5 | 18 | 70233555,917 | | |
| | | Total | 2124325062,6 | 19 | | | |
| 1,00 | 1 | Regression | 12814543052 | 1 | 12814543052,397 | 11,913 | ,003[a] |
| | | Residual | 19361576104 | 18 | 1075643116,897 | | |
| | | Total | 32176119157 | 19 | | | |

a. Predictors: (Constant), suprafata cultivata(ha)

b. Dependent Variable: valoarea serviciilor agricole(mil lei)

**Figura nr. 8.** *Tabelul ANOVA pentru cele două regresii*

- Se calculează valoarea testului Fisher. Deoarece valorile estimate sunt $RSS_1 = 1264204006{,}5$ şi $RSS_2 = 19361576104$, valoarea testului se calculează $F = \frac{RSS_2}{RSS_1}$.

Se obține $F= 1,53$, valoare pe care o comparăm cu valoarea teoretică $F_{\alpha;n_1-k;n_2-k} = F_{0,05;18;18} \approx 2,191$. Observăm că valoarea calculată este mai mică decât valoarea teoretică, deci acceptăm ipoteza $H_0$, adică modelul este homoscedastic.

### c. Ipoteza de necorelare a erorilor

Testarea acestei ipoteze se realizează cu ajutorul testelor: Runs Test și Durbin Watson test (DW).

**Runs Test** se bazează pe ideea că valorile variabilelor reziduale pot fi analizate ca un șir de secvențe de valori pozitive și negative. În ipoteza de necorelare a erorilor, acest șir este aleator sau numărul de secvențe (run-uri) este aleator și normal distribuit.

În SPSS, se urmează demersul:

*Analyze* → comanda *Nonparametric Tests* → opțiunea *Runs Test*.

Testul se face asupra variabilei „res_1".

*Observație*

În SPSS, secvențele de valori sau de run-uri se obțin prin compararea valorilor variabilei reziduale cu o valoare centrală (mediana, modul, media aritmetică) sau o altă valoare (figura 9). Run-urile se obțin din seturi de valori mai mari sau mai mici decât valoarea *cut point*.

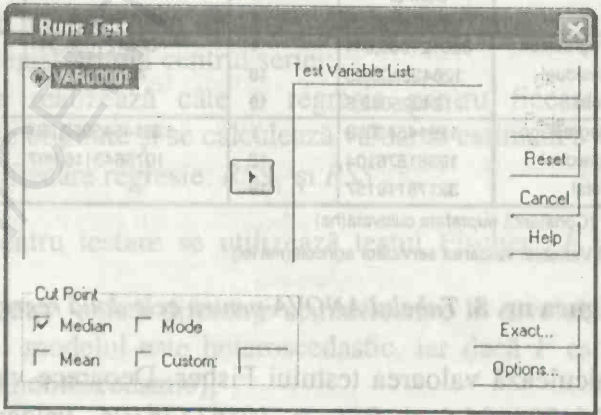



**Figura nr. 9.** *Runs-test*

BCU IASI CENTRAL UNIVERSITY LIBRARY

După ce în zona *Cut Point* se bifează opțiunea *Median*, se obține output-ul următor:

**Runs Test**

| | Unstandardized Residual |
|---|---|
| Test Value[a] | 616,52818 |
| Cases < Test Value | 20 |
| Cases >= Test Value | 20 |
| Total Cases | 40 |
| Number of Runs | 20 |
| Z | -,160 |
| Asymp. Sig. (2-tailed) | ,873 |

a. Median

**Figura nr. 10.** *Rezultate Runs-test*

Valoarea $sig = 0,873 > 0,05$ conduce la concluzia de a accepta ipoteza $H_0$, adică nu există fenomenul de autocorelare a erorilor.

Observăm, de asemenea, că valoarea testului $z = -0,16 \in (-1,96;1,96)$ se află în regiunea de acceptare $(-z_{\alpha/2}; z_{\alpha/2})$, de unde rezultă aceeași concluzie, de a accepta ipoteza nulă $(z_{\alpha/2} = z_{0,05/2} = z_{0,025} = 1,96)$.

**Durbin Watson Test (DW)** admite o ipoteză asupra coeficientului de corelație a erorilor ($\rho$). Ipotezele testului sunt:

$H_0: \rho = 0$

$H_1: \rho \neq 0.$

Demersul testării:

- Se calculează statistica $DW = d = 2(1-\hat{\rho})$.
- Regula de decizie:

• dacă $d \in [d_L, d_U] \cup [4-d_L, 4-d_U]$, nu se poate lua nici o decizie;

• dacă $d \in [d_U, 4-d_U]$, se acceptă $H_0$, nu există autocorelare;

• dacă $d \in [0, d_L]$, se respinge $H_0$, există autocorelare pozitivă;

• dacă $d \in [4-d_L, 4]$, se respinge $H_0$, exista autocorelare negativă.

BCU "LUCIAN BLAGA" CENTRAL UNIVERSITY LIBRARY

Unde: $d_L$ şi $d_U$ se citesc din tabele Durbin Watson.

În SPSS, se obţine valoarea testului $DW = d = 1{,}526$, utilizând opţiunea *Durbin-Watson*, în meniul ***Analyze*** → comanda ***Linear Regression***→ opţiunea ***Statistics***, conform figurii de mai jos:

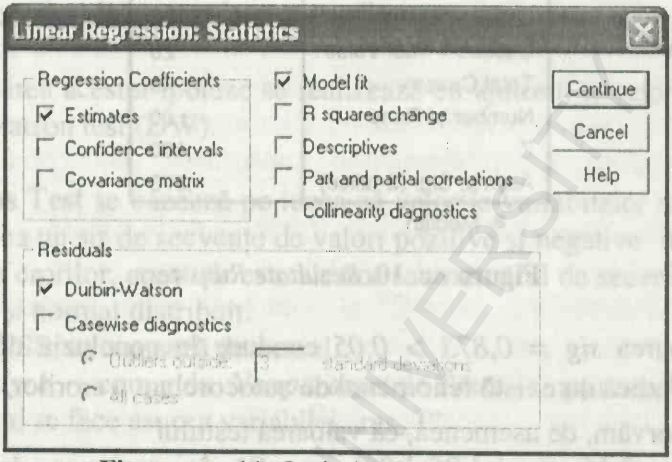



**Figura nr. 11.** Opţiunea *Durbin-Watson*

Din tabela Durbin Watson, pentru $k = 2$ (numărul de variabile independente), $n = 40$ (volumul eşantionului) şi un nivel de încredere de *95%*, se obţin valorile:

$d_L = 1{,}39$;

$d_U = 1{,}60$.

În concluzie, $d \in [d_L, d_U]$. Conform acestui test, nu se poate lua o decizie cu privire la această ipoteză.

### d. Ipoteza lipsei de coliniaritate a variabilelor independente

În SPSS, testarea acestei ipoteze se realizează cu ajutorul opţiunii de *Collinearity diagnostics* din meniul ***Analyze*** → comanda ***Linear Regression***→ optiunea ***Statistics***.

BCU "ION CREANGĂ" UNIVERSITARY LIBRARY

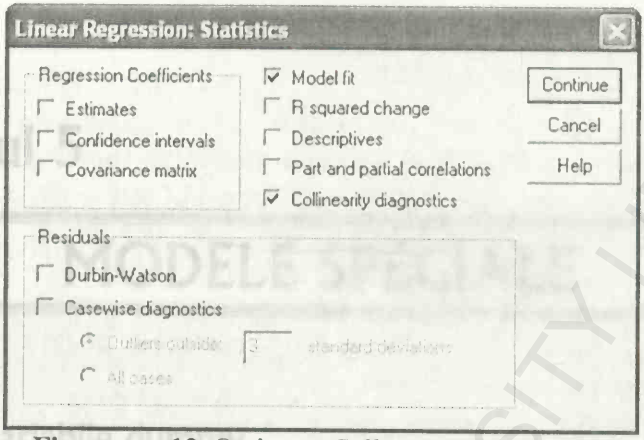



**Figura nr. 12.** Opțiunea *Collinearity diagnostics*

Rezultatul diagnosticului cu privire la ipoteza de necoliniaritate are la bază analiza a doi indicatori: *TOL* și *VIF*.

Dacă *VIF* are valori mai mari decât 10, există probleme de coliniaritate cu acea variabilă. În cazul în care *TOL = 0* există o coliniaritate perfectă; cu cât valorile *TOL* sunt mai aproape de 1, rezultă ca acea variabilă nu induce fenomenul de coliniaritate.

Rezultatele diagnosticului sunt prezente în figura 13.

**Coefficients[a]**

| | | Model | | |
|---|---|---|---|---|
| | | 1 | | |
| | | (Constant) | suprafata cultivata(ha) | nr mediu al salariatilor in agricultura(mii persoane) |
| Unstandardized Coefficients | B | -13515,86 | ,163 | 8,080 |
| | Std. Error | 6798,191 | ,047 | 2,132 |
| Standardized Coefficients | Beta | | ,447 | ,483 |
| t | | -1,988 | 3,502 | 3,790 |
| Sig. | | ,054 | ,001 | ,001 |
| Collinearity Statistics | Tolerance | | ,376 | ,376 |
| | VIF | | 2,659 | 2,659 |

a. Dependent Variable: valoarea serviciilor agricole(mil lei)

**Figura nr. 13.** Rezultatele pentru *Collinearity diagnostics*

Valorile indicatorilor *VIF* și *TOL* arată că nu sunt probleme în ceea ce privește coliniaritatea variabilelor independente.

BCU IAȘI/CENTRAL UNIVERSITY LIBRARY

# Capitolul 5

# MODELE SPECIALE

1. Variabile dummy
2. Modele ANOVA
3. Modele ANCOVA
4. Aplicaţii ale modelelor ANCOVA

Un domeniu mai puţin cercetat, adesea trecut uşor cu vederea, este cel al analizei de regresie cu variabile dummy. De regulă, în analiza de regresie, atât variabila rezultativă, cât şi variabila / variabilele factoriale sunt variabile numerice. Variabilele dummy sunt utilizate în calitate de variabile factoriale în analiza variaţiei prin procedeul ANOVA. Folosirea lor în analiza de regresie devine posibilă doar prin „cuantificarea" lor.

În lucrare, se consideră cazul analizei de regresie cu o variabilă rezultativă numerică şi una, respectiv, două variabile factoriale dummy.

BCU INSTITUTUL CENTRAL UNIVERSITY LIBRARY

## 5.1. Variabile dummy

Variabilele dummy sunt variabile calitative (nominale) care indică prezența sau absența unui atribut. De regulă, astfel de variabile sunt cunoscute în literatura de specialitate sub denumirea de variabile dichotomice, variabile binare sau variabile alternative. De exemplu, variabila „sex" prezintă două variante posibile, masculin – feminin. O persoană poate sau nu să posede o anumită categorie, poate sau nu fi de sex masculin, respectiv de sex feminin.

Generalizând, se poate construi o variabilă $X$ ale cărei valori sunt „da", când unitățile populației posedă categoria considerată, și „nu", când unitățile populației nu posedă acea categorie.

Fiecărei modalități i se poate atribui un cod, numeric sau simbolic, cu rol de identificator, care nu exprimă un raport de dependență între variantele variabilei. Fiecărei variante a variabilei $X$ i se atribuie o valoare, și anume: „1" pentru „da" și „0" pentru „nu". Sub această formă „cuantificabilă", variabilele dummy pot fi folosite în analiza de regresie. Sub formă cuantificabilă, variabila nominală dichotomică ia aspectul unei variabile numerice, tratamentul ei statistic devenind facil.

## 5.2 Modele ANOVA

Un model de regresie care conține doar variabile dummy ca variabile factoriale este cunoscut sub denumirea de ANOVA – model de analiză a variației.

Dacă variabilele dummy sunt „cuantificate", acestea pot fi incluse într-un model de regresie și pot fi tratate asemeni variabilelor cantitative. Un astfel de model de regresie poate fi scris sub forma modelului de tip liniar:

$$Y = \alpha + \beta \cdot D + \varepsilon$$

unde:

$Y$ = variabila rezultativă, numerică;

$D$ = variabila factorială, dummy, ale cărei valori sunt „da" și „nu", respectiv, „1" și „0";

BCU IAȘI CENTRAL UNIVERSITY LIBRARY

$\alpha$ = parametru, ordonata la origine a dreptei de regresie, care arată valoarea variabilei $Y$ când D=0;

$\beta$ = parametru, panta dreptei de regresie, numit coeficient de regresie;

$\varepsilon$ = reziduu sau variabila eroare aleatoare neobservabilă, de medie egală cu zero. Arată influenţa factorilor aleatori asupra variaţiei variabilei $Y$.

În modelul de regresie ANOVA, care conţine o variabilă dummy în calitate de variabilă factorială, media variabilei rezultative numerice $Y$ poate fi:

$$\left.\begin{aligned} M(Y \mid D_i = 0) &= \alpha \\ M(Y \mid D_i = 1) &= \alpha + \beta \end{aligned}\right\}$$

Ecuaţiile de mai sus arată că valoarea medie a variabilei $Y$ este egală cu valoarea parametrului $\alpha$, atunci când variabila dummy nu posedă însuşirea considerată; respectiv, cu suma celor doi parametri, $\alpha + \beta$, când variabila dummy posedă însuşirea. În cazul al doilea, valoarea parametrului $\beta$ arată cu cât este mai mare valoarea medie a variabilei $Y$, atunci când unităţile colectivităţii poartă însuşirea dată, faţă de cazul când nu poartă însuşirea dată.

În situaţia în care nu există nici o deosebire între valoarea medie a variabilei $Y$, obţinută pentru variabila dummy când $D = 0$ şi valoarea medie a variabile Y când $D = 1$, variabila dummy nu exercită o influenţă semnificativă asupra variabilei $Y$.

*Observaţie*

Dacă prin modelul de regresie se doreşte evidenţierea a două sau mai multe proprietăţi, se vor construi mai multe variabile dummy. De exemplu, dacă o variabilă nominală are $m$ variante, prin modelul ANOVA se poate evidenţia fiecare categorie construind $m-1$ variabile alternative. De exemplu, pentru o variabilă care admite $m=3$ valori ($V_1$ –$V_3$) se construiesc două variabile dummy, după cum urmează:

| | $D_1$ | $D_2$ |
|---|---|---|
| $V_1$ | 1 | 0 |
| $V_2$ | 0 | 1 |
| $V_3$ | 0 | 0 |

BCU IAŞI CENTRAL UNIVERSITY LIBRARY

În acest caz, modelul de regresie va fi:

$$Y = \alpha + \beta_1 D_1 + \beta_2 D_2 + \varepsilon.$$

Regresiile sau mediile condiționate au relațiile:

$$M(Y/D) = \begin{cases} \alpha, & D_{1i} = 0, D_{2i} = 0 \\ \alpha + \beta_1, & D_{1i} = 1, D_{2i} = 0 \\ \alpha + \beta_2, & D_{1i} = 0, D_{2i} = 1 \end{cases}$$

În esență, prin modelul de regresie ANOVA se rezolvă o problemă de inferență statistică clasică, și anume testarea existenței diferenței dintre două sau mai multe medii ale unor grupe de unități dintr-o populație distribuită după o variabilă *Y*.

Astfel, dacă se consideră doar două grupe și modelul de regresie $Y = \alpha + \beta D + \varepsilon$, se poate demonstra ușor că testarea parametrului $\beta$ din model este echivalentă cu testarea egalității mediilor celor două categorii de populație determinate de variabila dummy.

Notăm cu $\mu_1$ media variabilei dependente pentru prima categorie de unități din populație ($D_i = 1$) și cu $\mu_2$ media variabilei dependente pentru a doua categorie de unități din populație ($D_i = 0$).

Estimarea parametrilor modelului de regresie ANOVA se realizează prin metoda celor mai mici pătrate. Estimatorii parametrilor sunt:

$$\hat{\beta} = \hat{\mu}_1 - \hat{\mu}_2$$

$$\hat{\alpha} = \hat{\mu}_2$$

$$\hat{\alpha} + \hat{\beta} = \hat{\mu}_1$$

Testarea ipotezei $H_0$: $\beta = 0$ este echivalentă cu testarea ipotezei: $H_0$: $\mu_1 = \mu_2$.

## Exemplu

Pentru a exemplifica, considerăm variabilele din baza de date Tapestry:

BCU IASI CENTRALA UNIVERSITARA LIBRARY

- „sexul" (sexul persoanelor sosite în pelerinaj la Iaşi în perioada Sărbătorilor Iaşiului şi ale Sfintei Cuvioase Parascheva), variabilă independentă, dummy, cu variantele: 1 – Masculin, 0 – Feminin;

- „venit_l" (Venitul lunar al persoanelor sosite în pelerinaj la Iaşi în perioada Sărbătorilor Iaşiului şi ale Sfintei Cuvioase Parascheva), variabilă dependentă, numerică, continuă.

Pentru estimarea ecuaţiei de regresie am folosit programul SPSS, parcurgând paşii prezentaţi în figura 1, şi anume: meniul principal *Analyse* → comanda *Regression* → opţiunea *Linear*. Din fereastra de dialog *Linear Regression* se alege, din lista variabilelor, variabila „sexul" şi se mută în zona Variabilă independentă, iar variabila „venit_l" în zona Variabilă dependentă. Prin click asupra butonului OK se comandă output-ul, prezentat figura 2.

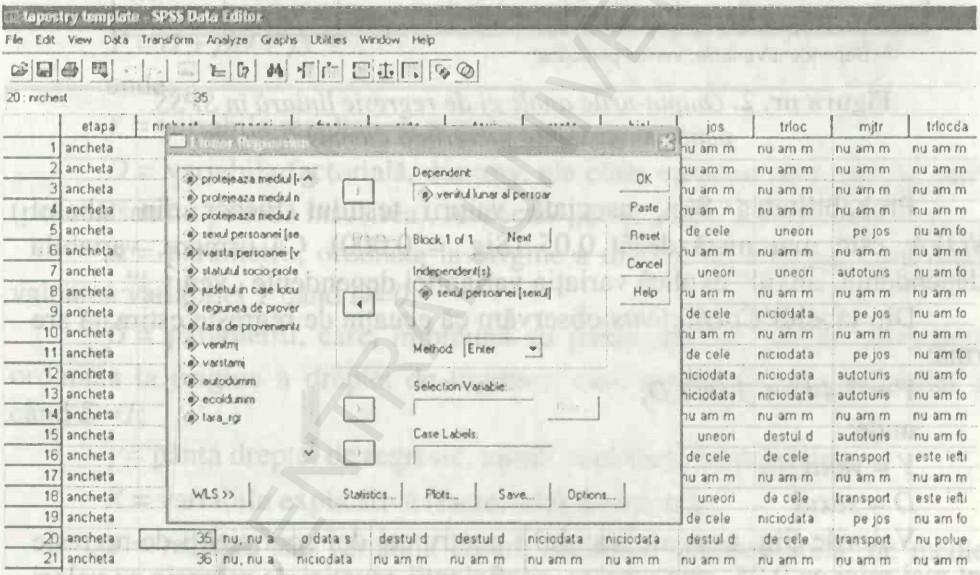



**Figura nr. 1.** *Fereastra de dialog Linear Regression în SPSS, pentru o variabilă factorială dummy*

ANOVA[b]

| Model | | Sum of Squares | df | Mean Square | F | Sig. |
|---|---|---|---|---|---|---|
| 1 | Regression | 98,451 | 1 | 98,451 | 15,962 | ,000[a] |
| | Residual | 2454,789 | 398 | 6,168 | | |
| | Total | 2553,240 | 399 | | | |

a. Predictors: (Constant), sexul persoanei

b. Dependent Variable: Venitul persoanei

Coefficients[a]

| Model | | Unstandardized Coefficients | | Standardized Coefficients | t | Sig. |
|---|---|---|---|---|---|---|
| | | B | Std. Error | Beta | | |
| 1 | (Constant) | 2,443 | ,164 | | 14,921 | ,000 |
| | sexul persoanei | 1,004 | ,251 | ,196 | 3,995 | ,000 |

a. Dependent Variable: Venitul persoanei

**Figura nr. 2.** *Output-urile analizei de regresie liniară în SPSS, pentru o variabilă factorială dummy*

Probabilitatea Sig., asociată valorii testului Fisher din tabelul *ANOVA*, este mai mică decât 0,05 (Sig. = 0,000). Ca urmare, variabila independentă „sexul" explică variația variabilei dependente „venit_l".

Din tabelul *Coefficients* observăm că ecuația de regresie estimată are forma:

$Y = 2,443 + 1,004\ D,$

unde:

$Y = venit\_l;$

$D = sexul$

Valorile Sig. asociate testului t pentru cei doi coeficienți de regresie sunt mai mici ca 0,05, prin urmare, coeficienții ecuației de regresie au valori semnificative la nivelul populației de pelerini. Valorile semnificative ale celor doi coeficienți de regresie arată că venitul lunar mediu este de 2,443 milioane, când variabila *„sexul persoanei"* ia valoarea *feminin*, respectiv de 3,447 milioane lei, când variabila independentă ia valoarea *masculin*, venitul mediu lunar este, deci cu 1,004 milioane lei mai mare decât în cazul în care variabila ia valoarea *feminin*.

BCU CENTRU UNIVERSITY LIBRARY

## 5.3. Modele ANCOVA

Un model de regresie care conţine o variabilă rezultativă numerică şi o mixtură de variabile cantitative şi variabile dummy ca variabile factoriale este cunoscut sub denumirea de model ANCOVA – model de analiză a covarianţei.

### a. Modele ANCOVA cu o variabilă dummy şi o variabilă numerică

Un model ANCOVA de acest tip poate fi scris sub forma modelului de tip liniar:

$$Y = \alpha + \beta D + \gamma X + \varepsilon$$

unde:

$Y$ = variabila rezultativă, numerică;

$D$ = variabila factorială, dummy, ale cărei variante sunt „da" şi „nu" (posedă sau nu posedă însuşirea), respectiv „1" şi „0";

$\alpha$ = parametru, ordonata la origine a dreptei de regresie, care arată valoarea variabilei $Y$ când $D=0$;

$\beta$ = parametru, care, împreună cu parametrul $\alpha$, $(\alpha+\beta)$, reprezintă ordonata la origine a dreptei de regresie, care arată valoarea variabilei $Y$ când $D=1$;

$\gamma$ = panta dreptei de regresie, numit coeficient de regresie;

$X$ = variabila explicativă (factorială), numerică;

$\varepsilon$ = reziduu sau variabila eroare aleatoare neobservabilă, de medie egală cu zero. Arată influenţa factorilor aleatori asupra variaţiei variabilei $Y$.

În cazul unui astfel de model, media variabilei rezultative Y, pentru fiecare din cele două valori ale variabilei factoriale dummy, poate fi:

$$\begin{cases} M(Y \mid X, D_i = 0) = \alpha + \gamma X \\ M(Y \mid X, D_i = 1) = (\alpha + \beta) + \gamma X \end{cases}$$

Din relaţiile de mai sus se desprinde concluzia că media variabilei rezultative $Y$, indiferent de valorile variantelor variabilei factoriale dummy,

BCU IASI / CENTRAL UNIVERSITY LIBRARY

este funcţie liniară de variabila explicativă cantitativă. Diferenţa dintre valoarea medie a variabilei rezultative în cazul în care elementele colectivităţii poartă însuşirea (deci, când $D = 1$), faţă de cazul în care elementele nu poartă însuşirea dată (deci, când $D = 0$) este egală cu $b$.

### Exemplu

Din baza de date Tapestry, considerăm variabilele:

- *„sexul"* (Sexul persoanelor sosite în pelerinaj la Iaşi în perioada Sărbătorilor Iaşiului şi ale Sfintei Cuvioase Parascheva), variabilă independentă, dummy, cu variantele: 1 – Masculin, 0 – Feminin;

- *„vârsta"* (Vârsta persoanelor sosite în pelerinaj la Iaşi în perioada Sărbătorilor Iaşiului şi ale Sfintei Cuvioase Parascheva), variabilă independentă, numerică, discretă;

- *„venit_l"* (Venitul lunar al persoanelor sosite în pelerinaj la Iaşi în perioada Sărbătorilor Iaşiului şi ale Sfintei Cuvioase Parascheva), variabilă dependentă, numerică, continuă.

Pentru estimarea ecuaţiei de regresie am folosit programul SPSS, parcurgând paşii prezentaţi în figura 2 şi anume: meniul principal *Analyse* → comanda *Regression* → opţiunea *Linnear*. Din fereastra de dialog *Linear Regression* se alege, din lista variabilelor, variabila „sexul" şi variabila „vârsta" ca variabile independente, iar variabila „venit_l" ca variabilă dependentă. Prin click asupra butonului OK se comandă output-ul, prezentat figura 4.

BCU IAŞI CENTRU UNIVERSITAR LIBRARY

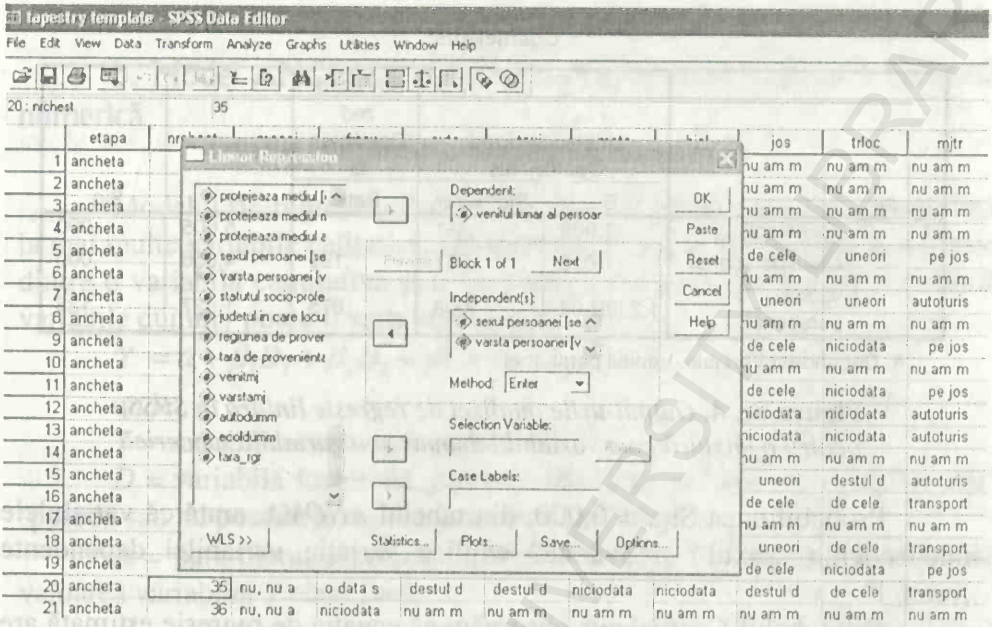



**Figura nr. 3.** *Fereastra de dialog Linear Regression în SPSS, pentru o mixtură cu o variabilă dummy și o variabilă numerică*

**ANOVA**[b]

| Model | | Sum of Squares | df | Mean Square | F | Sig. |
|---|---|---|---|---|---|---|
| 1 | Regression | 112,779 | 2 | 56,390 | 9,173 | ,000[a] |
| | Residual | 2440,461 | 397 | 6,147 | | |
| | Total | 2553,240 | 399 | | | |

a. Predictors: (Constant), varsta persoanei, sexul persoanei

b. Dependent Variable: Venitul persoanei

Coefficients[a]

| Model | | Unstandardized Coefficients | | Standardized Coefficients | t | Sig. |
|---|---|---|---|---|---|---|
| | | B | Std. Error | Beta | | |
| 1 | (Constant) | 1,969 | ,351 | | 5,605 | ,000 |
| | sexul persoanei | 1,018 | ,251 | ,199 | 4,056 | ,000 |
| | varsta persoanei | 1,239E-02 | ,008 | ,075 | 1,527 | ,128 |

a. Dependent Variable: Venitul persoanei

**Figura nr. 4.** *Output-urile analizei de regresie liniară în SPSS, pentru o mixtură cu o variabilă dummy şi o variabilă numerică*

Probabilitatea Sig. = 0,000, din tabelul *ANOVA*, arată că variabilele independente „sexul" şi „vârsta" explică variaţia variabilei dependente „venit_l".

Din tabelul *Coefficients* observăm că ecuaţia de regresie estimată are forma:

$Y = 1{,}969 + 1{,}018\,D + 0{,}01239\,X,$

unde:

$Y = venit\_l;$

$D = sexul\ persoanei;$

$X = vârsta.$

Valorile Sig., asociate testului t, arată că valorile coeficienţilor $a$ şi $b_1$ sunt semnificative, pe când coeficientul $b_2$ nu are o valoare semnificativă la nivelul populaţiei pelerinilor.

Valorile coeficienţilor de regresie arată că venitul lunar mediu, atunci când variabila *„sexul persoanei"* ia valoarea *feminin*, este cu 1,018 milioane lei mai mic decât atunci când variabila ia valoarea *masculin*. Totodată, la o creştere cu un an a vârstei unui pelerin, venitului lunar creşte, în medie, cu 0,01239 milioane. Valoarea mică a coeficientului de regresie $b_2$ ne arată că, pentru cazul considerat, vârsta nu este explicativă pentru creşterea venitului.

### b. Model ANCOVA cu două variabile dummy şi cu o variabilă numerică

Modelul de regresie cu variabile dummy poate fi extins cu uşurinţă la mai multe variabile calitative. Un model care exprimă legătura de regresie dintre o variabilă cantitativă şi o mixtură cu o variabilă cantitativă şi două variabile dummy poate fi scris sub forma:

$$Y = \alpha + \beta_1 D_1 + \beta_2 D_2 + \gamma X + \varepsilon$$

unde:

$Y$ = variabila rezultativă, numerică;

$D$ = variabila factorială, dummy, ale cărei variante sunt „da" şi „nu" (posedă sau nu posedă însuşirea), respectiv „1" şi „0";

$\alpha$ = parametru, ordonata la origine a dreptei de regresie, care arată valoarea variabilei $Y$ când D=0;

$\beta$ = parametru, care, împreună cu parametrul $\alpha$, $(\alpha+\beta)$, reprezintă ordonata la origine a dreptei de regresie, care arată valoarea variabilei $Y$ când D=1;

$\gamma$= panta dreptei de regresie, numit coeficient de regresie;

$X$ = variabila explicativă (factorială), numerică;

$\varepsilon$ = reziduu sau variabila eroare aleatoare neobservabilă, de medie egală cu zero. Arată influenţa factorilor aleatori asupra variaţiei variabilei $Y$.

Presupunând că $M(\varepsilon)=0$, pe baza relaţiei modelului, se pot obţine următoarele valori pentru media variabilei rezultative:

$$M(Y|X, D_1=0, D_2=0)=\alpha+\gamma X$$

$$M(Y|X, D_1=1, D_2=0)=(\alpha+\beta_1)+\gamma X$$

$$M(Y|X, D_1=0, D_2=1)=(\alpha+\beta_2)+\gamma X$$

$$M(Y|X, D_1=1, D_2=1)=(\alpha+\beta_1+\beta_2)+\gamma X$$

Se observă din aceste relaţii că ecuaţiile de regresie anterioare diferă numai prin valoarea parametrilor (a coeficienţilor variabilelor dummy), în timp ce panta dreptei de regresie $(\gamma)$ este aceeaşi.

O estimare cu ajutorul metodei celor mai mici pătrate a modelului ANCOVA va conduce la testarea mai multor ipoteze. Astfel, dacă $\beta_1$ este semnificativ din punct de vedere statistic, se poate desprinde concluzia că variabila $D_1$ influenţează în mare măsură variabila rezultativă $Y$. În mod

analog, dacă $\beta_2$ este statistic semnificativ, atunci variabila $D_2$ exercită o influenţă deosebită asupra variabilei rezultative. În cazul în care ambii parametri, $\beta_1$ şi $\beta_2$, sunt statistic semnificativi, se poate concluziona că cele două variabile dummy se constituie în factori determinanţi ai variabilei rezultative.

Din cele prezentate anterior, se poate observa faptul că un model poate fi extins prin includerea mai multor variabile cantitative (mai mult decât o singură variabilă cantitativă) şi a mai mult de două variabile calitative.

## Exemplu

Din baza de date Tapestry considerăm, în plus faţă de cazul precedent, variabila „ţara" (Ţara de proveniență a persoanelor sosite în pelerinaj la Iaşi în perioada Sărbătorilor Iaşiului şi ale Sfintei Cuvioase Parascheva), variabilă independentă, dummy, cu variantele: 1 – România, 0 – Alte ţări.

Output-ul rezultat în urma prelucrării este prezentat în figura 5.

**ANOVA[b]**

| Model | | Sum of Squares | df | Mean Square | F | Sig. |
|---|---|---|---|---|---|---|
| 1 | Regression | 189,206 | 3 | 63,069 | 10,565 | ,000[a] |
| | Residual | 2364,034 | 396 | 5,970 | | |
| | Total | 2553,240 | 399 | | | |

a. Predictors: (Constant), varsta, sexul persoanei, tara

b. Dependent Variable: Venitul persoanei

**Coefficients[a]**

| Model | | Unstandardized Coefficients | | Standardized Coefficients | t | Sig. |
|---|---|---|---|---|---|---|
| | | B | Std. Error | Beta | | |
| 1 | (Constant) | 4,246 | ,725 | | 5,861 | ,000 |
| | sexul persoanei | 1,010 | ,247 | ,198 | 4,083 | ,000 |
| | tara | -2,482 | ,694 | -,174 | -3,578 | ,000 |
| | varsta | 1,577E-02 | ,008 | ,095 | 1,958 | ,051 |

a. Dependent Variable: Venitul persoanei

**Figura nr. 5.** *Output-ul analizei de regresie liniară în SPSS, pentru o mixtură cu două variabile dummy şi o variabilă numerică*

Probabilitatea Sig. = 0,000, din tabelul *ANOVA*, arată că variabilele independente „sexul persoanei", „țara" și „vârsta" explică variația variabilei dependente „venitul".

Din tabelul *Coefficients* observăm că ecuația de regresie estimată are forma:

$$Y = 4{,}246 + 1{,}010\ D_1 - 2{,}482\ D_2 + 0{,}01577\ X,$$

unde:

$Y$ = *venit_l;*

$D_1$ = *sexul persoanei;*

$D_2$ = *țara;*

$X$ = *vârsta.*

Valorile Sig., asociate testului t, arată că valorile coeficienților de regresie $a$, $b_1$ și $b_2$ sunt semnificative, pe când coeficientul $b_3$ nu are o valoare semnificativă la nivelul populației pelerinilor, ceea ce înseamnă că doar variabilele „sexul" și „țara" influențează semnificativ variabila „venit_l".

Din valorile coeficienților de regresie se poate observa că venitul lunar mediu este cu 1,010 milioane lei mai mare atunci când variabila *„sexul"* ia valoarea *masculin* decât atunci când variabila ia valoarea *feminin* și este cu 2,482 mai mare atunci când variabila *„țara"* ia valoarea *Alte țări* decât atunci când valoarea variabilei este *România*. Totodată, la o creștere cu un an a vârstei unui pelerin, venitului lunar va crește, în medie, cu 0,01577 milioane.

Valorile medii ale variabilei rezultative Y, "venit_l", corespunzătoare sunt:

- venitul mediu lunar al pelerinilor de sex feminin din România este de 4,246 milioane lei;
- venitul mediu lunar al pelerinilor de sex masculin din România este de 5,256 milioane lei;
- venitul mediu lunar al pelerinilor de sex feminin din alte țări decât România este de 6,728 milioane lei;
- venitul mediu al pelerinilor de sex masculin din alte țări decât România este de 7,738 milioane lei.

## 5.4. Aplicaţii ale modelelor ANCOVA

### a. Compararea a două modele de regresie

În practică, există deseori situaţia în care este important să se compare variaţia unei variabile dependente la nivelul a două grupe în cadrul unei populaţii, a două populaţii diferite sau pentru două perioade de timp diferite.

Dacă se dispune de date statistice pentru două eşantioane de volum $n_1$, respectiv $n_2$, se pot compara modelele de regresie care exprimă variaţia variabilei dependente pentru situaţiile prezentate mai sus.

Considerăm situaţia cea mai simplă, a două variabile între care există o dependenţă liniară. Cele două modele de regresie corespunzătoare iau forma:

$$Y = \gamma_1 + \lambda_1 X + \varepsilon_1,\ i = \overline{1,n_1}$$

$$Y = \gamma_2 + \lambda_2 X + \varepsilon_2,\ i = \overline{1,n_2}$$

Pentru compararea celor două modele de regresie se poate construi un model ANCOVA de forma:

$$Y = \alpha_1 + \alpha_2 D + \beta_1 X + \beta_2 (XD) + \varepsilon, \text{ unde}$$

$D_i = 0$, dacă $i = \overline{1,n_1}$

$D_i = 1$, dacă $i = \overline{1,n_2}$.

În acest caz, regresia are forma:

$$M(Y/X,D) = \begin{cases} \alpha_1 + \beta_1 X, & D = 0 \\ (\alpha_1 + \alpha_2) + (\beta_1 + \beta_2) X, & D = 1 \end{cases}$$

Testarea parametrilor modelului ANCOVA conduce la următoarele situaţii posibile:

- se acceptă ipotezele $H_0$: $\alpha_2=0$, $H_0$: $\beta_2=0$, adică $\gamma_1=\gamma_2$, $\lambda_1=\lambda_2$ sau între cele două modele de regresie nu există diferenţe semnificative;

- se acceptă ipoteza $H_0$: $\beta_2=0$ şi se respinge ipoteza $H_0$: $\alpha_2=0$, adică $\gamma_1\neq\gamma_2$, $\lambda_1=\lambda_2$ sau cele două regresii sunt paralele şi diferă doar prin ordonata la origine, având aceeaşi pantă;

- se acceptă ipoteza $H_0$: $\alpha_2=0$ şi se respinge ipoteza $H_0$: $\beta_2=0$, adică $\gamma_1=\gamma_2$, $\lambda_1\neq\lambda_2$ sau cele două regresii diferă prin panta dreptei, având doar aceeaşi ordonată la origine;

- se resping ipotezele $H_0$: $\alpha_2=0$, $H_0$: $\beta_2=0$, adică $\gamma_1\neq\gamma_2$, $\lambda_1\neq\lambda_2$ sau între cele două modele de regresie există diferenţe semnificative.

## b. Testarea ipotezei de homoscedasticitate a unui model de regresie

Modelul ANCOVA cu o variabilă dummy poate fi folosit în testarea ipotezei de homoscedasticitate în cazul unui model de regresie. Acest procedeu de testare este bazat pe testul Goldfeld-Quandt.

Construim o variabilă dummy care împarte seria de date în două părţi egale, în funcţie de o valoare critică $X^*$ a variabilei independente $X$, după omiterea unui număr de înregistrări din centrul seriei. Valoarea critică se stabileşte în funcţie de valorile variabilei independente şi este o valoare aproape de mediană sau poate coincide cu aceasta. Se obţin două seturi de date cu valori mici, respectiv, cu valori mari ale variabilei independente. Omiterea valorilor din centrul seriei empirice ordonate are scopul omogenizării celor două seturi de date.

Variabila dummy ia valorile:
$D=1$, dacă $X_i < X^*$
$D=0$, în rest.

Testarea ipotezei de homoscedasticitate a unui model de regresie prin procedeul ANCOVA este echivalentă cu testarea semnificaţiei parametrilor corespunzători variabile dummy din următorul model:

$$Y=\alpha_1+\alpha_2 D+\beta_1 X+\beta_2 D\cdot X+\varepsilon.$$

Pentru acest model, mediile condiţionate sunt de forma:

$$M(Y/X,D)=\begin{cases}\alpha_1+\beta_1 X\ , & D=0\\ (\alpha_1+\alpha_2)+(\beta_1+\beta_2)X\ , & D=1\end{cases}$$

Decizia cu privire la ipoteza de homoscedasticitate echivalează cu decizia asupra semnificaţiei parametrilor $\alpha_2$, $\beta_2$, pe baza unui test Student. Dacă parametrii sunt semnificativi, înseamnă că cele două modele de regresie diferă semnificativ, deci modelul de regresie de bază este heteroscedastic. Dacă parametrii sunt nesemnificativi statistic, modelul de bază este homoscedastic.

BCU IASI/CENTRAL UNIVERSITY LIBRARY

# Capitolul 6

# MODELE PENTRU SERII DE TIMP

1. Componentele unei serii de timp
2. Estimarea trendului
3. Ajustarea seriilor sezoniere
4. Metode moderne de analiză a seriilor de timp
5. Exemple în SPSS

O serie de timp reprezintă un set de observații, un set de valori pe care le ia o variabilă (cantitativă sau calitativă) la diferite momente sau intervale de timp.

Principalele probleme ale modelării seriilor de timp sunt: delimitarea componentelor seriei, descompunerea seriei pe componente, ajustarea seriei în funcție de componentele identificate, prognoza fenomenului pe baza seriei ajustate.

BCU "CAROL I" CENTRAL UNIVERSITY LIBRARY

## 6.1. Componentele unei serii de timp

Seriile cronologice pot fi descompuse în patru componente, fiecare exprimând un aspect particular al valorilor seriei, definind şi tipologia lor. Aceste componente sunt: *tendinţa, fluctuaţia ciclică, variaţia sezonieră şi variaţia aleatoare,* prezentate grafic în figura 1.

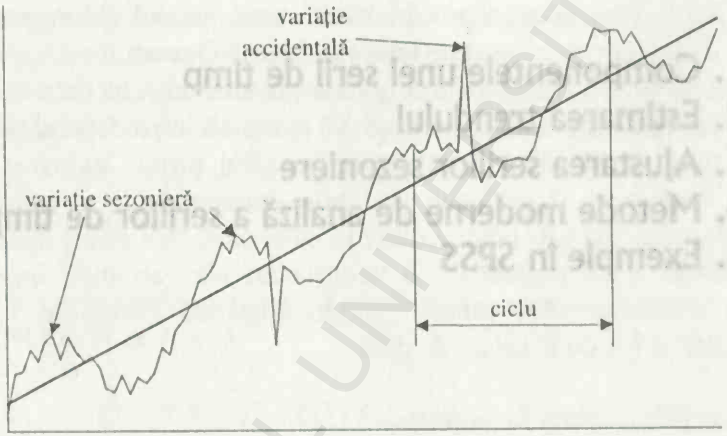



**Figura nr. 1.** *Componentele unei serii cronologice „complete"*

### Componenta tendenţială (trend)

*Componenta tendenţială,* numită şi *trend* sau *tendinţă seculară,* se notează $f_t$ şi *arată tendinţa generală înregistrată pe o perioadă lungă de timp şi exprimă variaţia medie,* respectiv, *legea de evoluţie* a variabilei observate. Este de natură deterministă, timpul fiind singura variabilă care explică variaţia fenomenului observat.

Descrierea componentei trend se face printr-o *funcţie continuă,* de regulă, printr-o funcţie polinomială în timp, $y_t = f(t)$, de grad $n$ (cu $n=1$ sau 2, foarte rar $n>2$). Aproximarea modelului de trend (a legii de evoluţie în timp a variabilei observate) şi estimarea parametrilor acestuia se constituie ca obiect al *ajustării statistice.*

## Componenta ciclică

Componenta ciclică se notează $C_t$ şi *se prezintă ca o fluctuaţie de tip sinusoidal, în jurul trendului*, cu o durată relativ lungă de timp, cu caracter de repetabilitate regulată. Variaţia ciclică cuprinde patru faze: expansiunea, criza, recesiunea şi relansarea (vezi figura 2.).

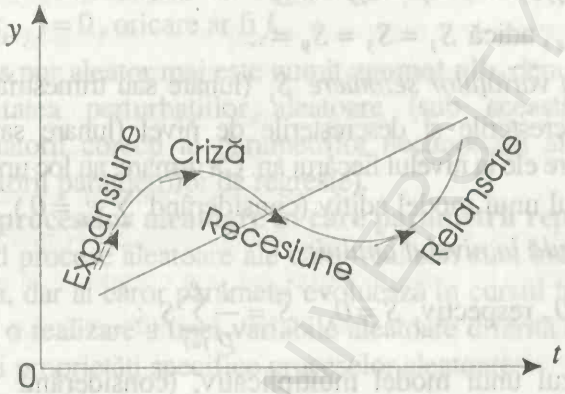



**Figura nr. 2.** *Fazele unui ciclu*

În activitatea economică, s-au conturat cicluri cu durată diferită, de exemplu, de aproximativ *50* ani pentru ciclul tip Kondratieff, de aproximativ *9* ani pentru ciclul tip Juglar, de aproximativ *7* ani pentru ciclul biblic (*7* vaci grase, *7* vaci slabe).

Analiza componentei ciclice (atunci când aceasta există, deoarece adesea se suprapune pe trend) presupune studierea comportamentului variabilei observate pe o perioadă lungă de timp în raport cu trendul, decupându-se periodicitatea ciclică şi fazele unui ciclu.

## Componenta sezonieră

Componenta sezonieră se notează $S_t$ şi este definită de *variaţiile pe termen scurt, prin creşteri şi descreşteri de nivel succesive care se repetă de la o perioadă la alta,* fie în jurul componentei ciclice (când aceasta există), fie în jurul componentei trend.

Variaţiile sezoniere sunt repetabile, de regulă, de la o lună la alta, sau de la un trimestru la altul. La nivelul unui an, influenţa sezonieră este neutră.

Teoretic, *variaţiile sezoniere,* $S_t$, *se repetă riguros identic de la o perioadă „p" la alta*, adică:

$$S_t = S_{t+p} = S_{t+2p} = ... = etc.$$

Dacă, de exemplu, perioadele ar fi lunile sau trimestrele anului, atunci variaţiile sezoniere corespunzătoare se repetă identic lunar sau trimestrial în fiecare an, respectiv, au loc egalităţile:

$S_t = S_{t+12}$, adică $S_1 = S_{13} = S_{25} = ...$ etc., sau

$S_t = S_{t+4}$, adică $S_1 = S_5 = S_9 = ...$

*Influenţa variaţiilor sezoniere* $S_t$ (lunare sau trimestriale) *este neutră la nivelul anului*; creşterile şi descreşterile de nivel (lunare sau trimestriale) se compensează între ele la nivelul fiecărui an. Ca urmare, au loc următoarele relaţii:

- în cazul unui model aditiv (considerând $\sum \varepsilon_t = 0$), *media variaţiilor sezoniere este nulă la nivelul anului:*

$$\sum_{t=1}^{p} S_t = 0, \text{ respectiv, } \overline{S} = 0, \quad \overline{S} = \frac{1}{p}\sum_{t=1}^{p} S_t;$$

- în cazul unui model multiplicativ, (considerând $\sum \varepsilon_t = 0$) *media variaţiilor sezoniere este egală cu unitatea:*

$$\overline{S} = 1, \overline{S} = \frac{1}{p}\sum_{t=1}^{p} S_t;$$

- în cazul modelului multiplicativ, variaţiile sezoniere cresc sau descresc cu aceeaşi *rată* (proporţie), *s*, astfel că:

$$\overline{S} = 1 + s,$$

unde $1+s$ este *multiplicatorul,* iar $\overline{S}$ *rata medie* de creştere, egală cu zero: $\overline{S} = 0$. Adică, analog modelului aditiv, rata de creştere a variaţiilor sezoniere se anulează, în medie, în fiecare an, repetându-se identic de la an la an.

Variaţiile sezoniere sunt datorate unor cauze diferite care definesc ritmul activităţilor sezoniere (periodicitatea lucrărilor agricole, a concediilor, a sărbătorilor tradiţionale). Depistarea componentei sezoniere se face prin metode grafice, armonice.

## Componenta aleatoare (reziduală)

Componenta aleatoare se notează $\varepsilon_t$ şi are drept caracteristică de bază *caracterul non-determinist al variaţiei.* Se consideră că pe un număr mic de ani $\varepsilon_t$ se compensează: $\sum \varepsilon_t = 0$. Variaţia aleatoare se poate manifesta ca *un*

*proces pur aleator, un proces aleator în care parametrii variază în timp, şi ca un proces staţionar.*

În **cazul procesului pur aleator,** seria cronologică este o secvenţă de variabile independente mutual şi cu aceeaşi distribuţie de probabilitate. Speranţa matematică şi varianţa variabilei aleatoare $\varepsilon_t$ sunt independente de timp:

$M(\varepsilon_t) = \mu \quad V(\varepsilon_t) = \sigma_\varepsilon^2$, oricare ar fi perioada *t,* iar covarianţa a două perturbaţii aleatoare decalate de un număr *j* de perioade este nulă:

$Cov(\varepsilon_t, \varepsilon_{t+j}) = 0$, oricare ar fi *j.*

Un proces pur aleator mai este numit **zgomot alb**, denumire ce implică necesar normalitatea perturbaţiilor aleatoare (sub această ipoteză s-au determinat estimatorii corecţi ai parametrilor necunoscuţi ai unei populaţii, precum şi estimatorii parametrilor de regresie).

În **cazul proceselor aleatoare în care parametrii repartiţiei variază în timp** se includ procese aleatoare ale căror realizări sunt independente de la o perioadă la alta, dar ai căror parametri evoluează în cursul timpului. Fiecare dată a seriei este o realizare a unei variabile aleatoare diferită de precedentele. Respectă aceleaşi proprietăţi specifice proceselor aleatoare:

$M(\varepsilon_t) = \mu; \quad V(\varepsilon_t) = \sigma_{\varepsilon_t}^2, \quad \text{cov}(\varepsilon_t, \varepsilon_{t+j}) = 0.$

În **cazul proceselor staţionare,** caracteristică este dependenţa variabilei $\varepsilon_t$ de $\varepsilon_{t-1}, \varepsilon_{t-2}, ..., \varepsilon_{t-k}$, dependenţă exprimată prin proprietatea: *covarianţa între* $\varepsilon_t$ *şi* $\varepsilon_{t-j}$ *rămâne constantă şi nu depinde decât de decalajul j şi nu de t.* Ca urmare, proprietăţile specifice proceselor staţionare, în raport cu procesele pur aleatoare, sunt:

$M(\varepsilon_t) = \mu \quad V(\varepsilon_t) = \sigma_\varepsilon^2$, pentru oricare *t,*

$\text{cov}(\varepsilon_t, \varepsilon_{t-j}) \neq 0$, rămâne constantă şi nu depinde decât de *j* şi nu de *t.*

## Combinarea componentelor unei serii cronologice

Componentele unei serii cronologice, $f_t$ trendul, $C_t$ componenta ciclică, $S_t$ componenta sezonieră, $\varepsilon_t$ componenta aleatoare, se pot combina fie aditiv, fie multiplicativ. Din combinările posibile, în practica economică se aplică frecvent următoarele modele:

- modelul aditiv: $y_t = f_t + C_t + S_t + \varepsilon_t$
- modelul multiplicativ: $y_t = f_t \cdot C_t \cdot S_t \cdot \varepsilon_t$

- modelul mixt: $y_t = S_t + (f_t \cdot C_t \cdot \varepsilon_t)$

sau $y_t = C_t + (f_t \cdot S_t \cdot \varepsilon_t)$.

În majoritatea cazurilor, modelul multiplicativ se transformă logaritmic: $y_t = f_t \cdot S_t \cdot \varepsilon_t$ este echivalent cu $\log y_t = \log f_t + \log S_t + \log \varepsilon_t$.

Transformarea se recomandă pentru a stabiliza varianța sa și pentru a facilita calculele.

Într-un model aditiv, fenomenul $y$ studiat în funcție de timp se descompune în componente independente unele de altele, pe când într-un model multiplicativ $y_t$ se descompune în componente dependente unele de altele.

Analiza unei serii cronologice constă în determinarea valorilor luate de fiecare din cele pentru componente. Tratarea componentelor impune să se înceapă cu evaluarea trendului ($f_t$), apoi să se estimeze variațiile sezoniere și, în fine, componenta ciclică (atunci când aceasta există). Ceea ce nu se atribuie celor trei componente se consideră explicat prin componenta aleatoare ($\varepsilon_t$). În următoarele paragrafe, vom insista asupra evaluării acestor componente, în special, asupra trendului.

## 6.2. Estimarea trendului

Metoda analitică de ajustare *constă în aproximarea tendinței de variație în timp cu ajutorul funcției de trend* $y_t = f(t)$, în care „*t*" reprezintă variabila timp, iar „y" variabila în timp. Prin folosirea metodei analitice de ajustare *se stabilește cu mai multă siguranță legea de dezvoltare*, de variație a unui fenomen, deoarece se ține seama de valorile empirice ale tuturor termenilor unei serii.

Tendința de variație a fenomenelor sociale se aproximează, de regulă, printr-una din funcțiile ale căror curbe și ecuații de estimare au fost prezentate. Variabila *x* din modelele de regresie devine *t* în modelele de trend.

Parametrii ecuației de trend se află, de regulă, cu ajutorul metodei celor mai mici pătrate.

**Trendul liniar.** O variație în timp de forma unei funcții liniare este specifică fenomenelor care se dezvoltă rectiliniu. Ecuația de estimare a funcției de trend liniare este definită de relația:

$y_t = a + bt$

Determinarea parametrilor funcției de trend are la bază **principiul celor mai mici pătrate,** conform căruia trebuie să se satisfacă condiția:

$$S = \sum e_i^2 = \sum (y_i - y_t)^2 = min\,im$$

Aflarea parametrilor presupune următoarele **operații**:

1. *se află derivatele parțiale ale expresiei de mai sus în raport cu parametrii funcției* $y_t = f(t)$;

2. *se anulează derivatele parțiale;*

3. *se rezolvă sistemul de ecuații normale obținut.*

Pentru un trend liniar, principiul celor mai mici pătrate cere să se satisfacă condiția:

$$S = \sum (y_i - a - bt_i)^2 = min\,im\,.$$

Se stabilesc și se anulează derivatele parțiale ale expresiei „*S*" în raport cu *a* și *b*:

$$\frac{\partial S}{\partial a} = 2\sum (y_i - a - b \cdot t_i)(-1) = 0$$

$$\frac{\partial S}{\partial b} = 2\sum (y_i - a - b \cdot t_i)(-t_i) = 0$$

De aici, rezultă următorul sistem de două ecuații normale:

$$n \cdot a + b\sum t_i = \sum y_i$$

$$a \cdot \sum t_i + b\sum t_i^2 = \sum t_i y_i$$

În funcția de trend, variabila timp reprezintă, de regulă, o serie de numere consecutive, originea luându-se în centrul seriei și deci $\sum t_i = 0$. Ca urmare, ecuațiile normale pot fi simplificate, luând următoarea formă:

$$n \cdot a = \sum y_i$$

$$b\sum t_i^2 = \sum t_i y_i$$

de unde rezultă relațiile de calcul pentru parametrii ecuației de trend:

$$a = \frac{\sum y_i}{n} \text{ și } b = \frac{\sum t_i y_i}{\sum t_i^2}$$

Valorile variabilei timp se măsoară în unități întregi (intervale), dacă numărul termenilor seriei este impar și anume:

*..., -2, -1, 0, 1, 2, ...*

sau în jumătăți de interval, dacă numărul termenilor este par, astfel:

*..., -5, -3, -1, 0, 1, 3, 5, ...*

BCU "MIHAI EMINESCU" CENTRAL UNIVERSITY LIBRARY

Aflarea parametrilor se efectuează mai facil dacă elementele de calcul sunt grupate într-un tabel (vezi tabelul 1).

**Exemplu.** Considerând datele prezentate în tabelul 1 cu privire la un fenomen *Y*, într-o perioadă de *11* ani, se cere să se ajusteze seria prin metoda analitică.

**Tabelul nr. 1.** *Elemente de calcul pentru determinarea tendinţei liniare*

| Anii | $y_i$ | $t_i$ | $t_i y_i$ | $t_i^2$ | Valorile seriei ajustate $y_{t_i} = a + bt_i = 56{,}9 + 9{,}44t_i$ |
|---|---|---|---|---|---|
| 1 | 2 | 3 | 4 | 5 | 6 |
| 1 | 22,9 | -5 | -114,5 | 25 | 14,7 |
| 2 | 25,2 | -4 | -180,8 | 16 | 23,14 |
| 3 | 28,8 | -3 | - 86,4 | 9 | 31,58 |
| 4 | 36,0 | -2 | - 72,8 | 4 | 40,82 |
| 5 | 49,8 | -1 | - 49,8 | 1 | 48,46 |
| 6 | 53,1 | 0 | 0 | 0 | 56,9 |
| 7 | 60,8 | 1 | 68,8 | 1 | 65,34 |
| 8 | 69,6 | 2 | 139,2 | 4 | 73,78 |
| 9 | 77,4 | 3 | 232,2 | 9 | 82,22 |
| 10 | 92,3 | 4 | 369,2 | 16 | 90,66 |
| 11 | 110,0 | 5 | 550,8 | 25 | 99,10 |
| Total | 625,9 | 0 | 927,9 | 110 | 625,9 |

Introducând datele, calculate în tabelul 1, în sistemul de ecuaţii se va obţine:

$11a = 625{,}9$

$110b = 927{,}9$

de unde rezultă valoarea parametrilor:

$a = 56{,}9$ şi $b = 8{,}44$

Ecuaţia tendinţei liniare va fi:

$y_t = a + bt = 56{,}9 + 8{,}44t$,

iar dreapta determinată pe baza acestei ecuaţii este prezentată în coloana 6 a tabelului 1.

Verificarea exactităţii ajustării constă în compararea sumei valorilor empirice cu suma valorilor ajustate (teoretice) ale termenilor seriei, care trebuie să fie egale ($\sum y_i = \sum y_t$), în ipoteza $\sum e_i = 0$.

În seria considerată, abaterile în plus şi în minus faţă de tendinţa medie sunt relativ mici, ceea ce ne îndreptăţeşte să credem că dreapta calculată oglindeşte corect trendul în perioada cercetată.

**Trendul parabolic.** Trendul parabolic este specific fenomenelor care prezintă o tendinţă crescătoare sau descrescătoare cu un punct de maxim, respectiv de minim.

Ecuaţia tendinţei parabolice de gradul doi este:

$$y_t = a + b \cdot t + c \cdot t^2$$

Aplicând metoda celor mai mici pătrate, în cazul ajustării după o parabolă de gradul doi, se obţine următorul sistem de ecuaţii normale:

$$n \cdot a + b \cdot \sum t_i + c \cdot \sum t_i^2 = \sum y_i$$
$$a \cdot \sum t_i + b \cdot \sum t_i^2 + c \cdot \sum t_i^3 = \sum t_i y_i$$
$$a \cdot \sum t_i^2 + b \cdot \sum t_i^3 + c \cdot \sum t_i^4 = \sum t_i^2 y_i$$

În condiţia $\sum t_i = 0$, se obţin:

$$a = \frac{\sum y_i \cdot \sum t_i^4 - \sum t_i^2 y_i \cdot \sum t_i^2}{n \cdot \sum t_i^4 - (\sum t_i^2)^2}$$

$$b = \frac{\sum t_i y_i}{\sum t_i^2}$$

$$c = \frac{n \cdot \sum t_i^2 y_i - \sum t_i^2 \cdot \sum y_i}{n \cdot \sum t_i^4 - (\sum t_i^2)^2}$$

### Exemplu

Considerăm datele din exemplul de mai sus. Să se ajusteze seria după un trend parabolic.

### Rezolvare

Elementele de calcul sunt prezentate în tabelul 2. Înlocuind datele din tabelul 2 în sistemul de ecuaţii, se obţine soluţia sistemului:

$$11a + 110c = 625{,}9 \qquad b = 8{,}435454 \approx 8{,}44$$
$$110b = 927{,}9 \qquad a = 51{,}861538 \approx 51{,}86$$
$$110a + 1958c = 6691{,}3 \qquad c = 0{,}503846 \approx 0{,}504$$

Ecuaţia estimată a tendinţei parabolice pentru seria dată va fi:

$y_t = a + bt + ct^2 = 51{,}86 + 8{,}44t + 0{,}504t^2$,

iar valorile teoretice pentru perioada 1-11 sunt prezentate în coloana a 8-a din tabelul 2.

**Tabelul nr. 2.** *Elemente de calcul*

| Anii | Seria empirică $y_i$ | $t_i$ | $t_i^2$ | $t_i^4$ | $t_i y_i$ | $t_i^2 y_i$ | Seria ajustată $y_{t_i} = a + bt_i + ct_i^2 = 51{,}86 + 8{,}44t_i + 0{,}504t_i^2$ |
|---|---|---|---|---|---|---|---|
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
| 1 | 22,9 | -5 | 25 | 625 | -114,5 | 572,5 | 22,26 |
| 2 | 25,2 | -4 | 16 | 256 | -100,8 | 403,2 | 26,164 |
| 3 | 28,0 | -3 | 9 | 81 | -86,4 | 259,2 | 31,076 |
| 4 | 36,0 | -2 | 4 | 16 | -72,0 | 144,0 | 36,996 |
| 5 | 49,8 | -1 | 1 | 1 | -49,8 | 49,8 | 43,924 |
| 6 | 53,1 | 0 | 0 | 0 | 0 | 0 | 51,06 |
| 7 | 60,8 | 1 | 1 | 1 | 60,8 | 60,8 | 68,804 |
| 8 | 69,6 | 2 | 4 | 16 | 139,2 | 278,4 | 70,756 |
| 9 | 77,4 | 3 | 9 | 81 | 232,2 | 696,6 | 81,716 |
| 10 | 92,3 | 4 | 16 | 256 | 369,2 | 1476,8 | 93,684 |
| 11 | 110,0 | 5 | 25 | 625 | 550,0 | 2750,0 | 106,66 |
| Total | 625,9 | 0 | 110 | 1958 | 927,9 | 6691,3 | 625,906 |

**Trendul exponențial.** Trendul exponențial este specific fenomenelor care se dezvoltă asemănător unei progresii geometrice.

Ecuația de estimare a trendului exponențial este dată de relația:

$y_t = a \cdot b^t$

Prin logaritmarea expresiei funcției de trend, se obține:

$\log y_t = \log a + t \log b$

Aplicând metoda celor mai mici pătrate în ajustarea după o curbă exponențială, a cărei ecuație a fost logaritmată și considerând condiția $\sum t_i = 0$, se obțin formulele de calcul ale parametrilor modelului de trend:

$$\log a = \frac{\sum \log y_i}{n} \text{ și } \log b = \frac{\sum t_i \log y_i}{\sum t_i^2}$$

## 6.3. Ajustarea seriilor sezoniere

Cunoaşterea variaţiilor sezoniere necesită depistarea componentei sezoniere (prin metode grafice), măsurarea lor (prin indici şi coeficienţi de sezonalitate) şi analiza diferitelor cauze care definesc ritmul producerii lor (de exemplu, periodicitatea lucrărilor agricole, a concediilor, a sărbătorilor etc.).

Ajustarea seriilor cronologice cu variaţii sezoniere se bazează pe descompunerea seriei, pe de o parte, în componenta trend ( $f_t$ ) şi componenta ciclică C (când este cazul) şi, pe de altă parte, în componenta variaţie sezonieră ( $S_t$ ), considerând influenţa aleatoare nulă ( $\sum \varepsilon_t = 0$ ), adică:

$y_t = f_t + S_t$ (pentru un model aditiv), respectiv

$y_t = f_t \cdot S_t$ (pentru un model multiplicativ).

Ajustarea componentei trend a fost tratată în paragraful anterior. Ajustarea componentei sezoniere *presupune eliminarea influenţei sezoniere.* Ajustarea seriilor sezoniere constă în *înlocuirea termenilor reali ai seriei cu termeni calculaţi şi are ca rezultat obţinerea unei serii cronologice cu variaţii sezoniere riguros identice de la o perioadă la alta şi cu influenţă nulă la nivelul fiecărui an.* Această operaţie se face, de regulă, prin metoda mediilor mobile. Prin această metodă se definesc componenta trend şi componenta ciclică. Pentru „capturarea" componentei sezoniere, se calculează indicii de sezonalitate. Pentru izolarea componentei sezoniere, se calculează coeficienţii sezonieri ($S_j$), pe baza cărora se desezonalizează seria iniţială. Această operaţie se face prin raportarea valorilor aberante ($y_i$) la coeficienţii sezonieri.

*Resezonalizarea* este operaţia inversă desezonalizării, aplicată asupra valorilor calculate prin ajustare. Resezonalizarea se realizează în funcţie de modelul de compunere admis, în sens previzional.

**Depistarea grafică a variaţiilor sezoniere.** Grafic, seria ajustată se prezintă sub forma unei linii ondulatorii, cu bucle riguros identice, mai aplatizate faţă de cronograma iniţială, repetabile în jurul liniei de trend.

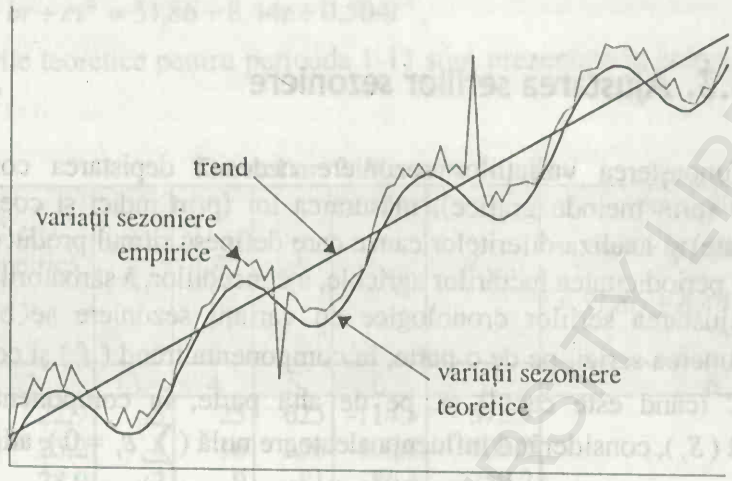



**Figura nr. 3.** *Trendul și variațiile sezoniere empirice și teoretice*

## Ajustarea prin medii mobile

Ajustarea seriilor cronologice cu variații sezoniere prin metoda mediilor mobile *constă în înlocuirea termenilor empirici cu termeni rezultați în urma calculării mediilor mobile pentru seria dată.* Prin înlocuirea termenilor reali cu termeni calculați, va rezulta o curbă mai rotunjită sau o dreaptă de tendință, cu condiția ca să se fi determinat corect periodicitatea de variație a fenomenului. Periodicitatea este evidențiată de punctele de maxim sau de minim (vezi fig. 3).

*Calculul mediilor mobile constă în aflarea mediilor aritmetice dintr-un număr impar sau par de termeni luați succesiv*, în funcție de mărimea unui ciclu de variație.

Când numărul de termeni luați în calcul este impar, mediile obținute cad pe termeni reali pe care-i vor înlocui. Când se cuprinde în calcul un număr par de termeni, mediile cad între doi termeni reali. Pentru a afla ce termen va fi înlocuit se face *centrarea mediilor mobile*, adică se determină media aritmetică din două medii mobile consecutive. Procedeul de determinare a mediilor mobile este evidențiat în tabelul 3.

După cum se observă și din tabelul 3, numărul termenilor din seria ajustată prin medii mobile este egal cu:

$N - (n - 2) - 1$, respectiv $N - (n - 2) - 2$,

unde: $N$ reprezintă numărul termenilor din seria empirică, iar
$n$, numărul termenilor cuprinși în calculul mediilor mobile.

Prima relație este pentru un $n$ impar, iar a doua, pentru un $n$ par. De exemplu, când $N = 7$, iar $n = 3$, respectiv $n = 4$, rezultă că seria ajustată poate avea $7-(3-2)-1=5$ termeni, respectiv $7 - (4 - 2) - 2 = 3$ termeni, caz ilustrat și în tabelul 3. Devierile față de valoarea medie, datorate sezonalității, pot fi măsurate prin *indici de sezonalitate*.

**Indicii de sezonalitate.** Indicii de sezonalitate se determină ca raport între termenii reali și cei ajustați, după relațiile:

$$i = \frac{y_i}{\bar{y}_i} \cdot 100, \text{ respectiv } i = \frac{y_i}{\bar{\bar{y}}_t} \cdot 100, \text{ sau } i = \frac{y_i}{\bar{y}_t},$$

când ajustarea s-a efectuat prin medii mobile, respectiv, prin metoda celor mai mici pătrate.

**Tabelul nr. 3.** *Elemente de calcul (ajustarea prin medii mobile)*

| Valori empirice | Medii mobile calculate din | | Medii centrate | Indici de sezonalitate | |
|---|---|---|---|---|---|
| | număr impar (de ex. trei) (valori ajustate) | număr par (de ex. patru) | | | |
| $y_1$ | - | - | - | - | - |
| $y_2$ | $\bar{y}_1 = \frac{y_1 + y_2 + y_3}{3}$ | $\bar{y}_1 = \frac{y_1 + y_2 + y_3 + y_4}{4}$ | $\bar{\bar{y}}_1 = \frac{\bar{y}_1 + \bar{y}_2}{2}$ | $\frac{y_2}{\bar{\bar{y}}_1} \cdot 100$ | - |
| $y_3$ | $\bar{y}_2 = \frac{y_2 + y_3 + y_4}{3}$ | $\bar{y}_2 = \frac{y_2 + y_3 + y_4 + y_5}{4}$ | $\bar{\bar{y}}_2 = \frac{\bar{y}_2 + \bar{y}_3}{2}$ | $\frac{y_3}{\bar{\bar{y}}_2} \cdot 100$ | $\frac{y_3}{\bar{y}_1} \cdot 100$ |
| $y_4$ | $\bar{y}_3 = \frac{y_3 + y_4 + y_5}{3}$ | $\bar{y}_3 = \frac{y_3 + y_4 + y_5 + y_6}{4}$ | $\bar{\bar{y}}_3 = \frac{\bar{y}_3 + \bar{y}_4}{2}$ | $\frac{y_4}{\bar{\bar{y}}_3} \cdot 100$ | $\frac{y_4}{\bar{\bar{y}}_2} \cdot 100$ |
| $y_5$ | $\bar{y}_4 = \frac{y_4 + y_5 + y_6}{3}$ | $\bar{y}_4 = \frac{y_4 + y_5 + y_6 + y_7}{4}$ | | $\frac{y_5}{\bar{y}_4} \cdot 100$ | $\frac{y_5}{\bar{\bar{y}}_3} \cdot 100$ |
| $y_6$ | $\bar{y}_5 = \frac{y_5 + y_6 + y_7}{3}$ | | | $\frac{y_6}{\bar{y}_5} \cdot 100$ | - |
| $y_7$ | | | | | |
| ... | | | | | |

## Coeficienţi sezonieri

Variaţiile sezoniere ($S_t$) se repetă, *teoretic*, identic de la o perioadă la alta (lună de lună, trimestru de trimestru) şi se compensează la nivelul anului, conform principiului de conservare a ariilor. Practic, variaţiile sezoniere nu se repetă absolut identic.

Pentru a ajusta o serie reală, respectând exigenţele modelului teoretic, variaţiile sezoniere $S_t$ observate se înlocuiesc cu valori calculate numite *coeficienţi sezonieri*, $S_j, j = \overline{1,p}$ perioade ($j = \overline{1,12}$, pentru luni, respectiv $j = \overline{1,4}$, pentru trimestre). *Coeficienţii sezonieri* $S_j$ *sunt identici perioadă de perioadă pe „n" ani observaţi*, adică există „*j*" coeficienţi sezonieri şi nu „*j* x *n*" variaţii sezoniere $S_t$ pe *n* ani. (De exemplu: *16* $S_t$ diferite pentru *4* ani, pe trimestru, sau *24* $S_t$ diferite pentru *2* ani, lunar).

**Calculul coeficienţilor sezonieri.** *Coeficienţii sezonieri se calculează ca o medie aritmetică a variaţiilor sezoniere*, lună de lună sau trimestru de trimestru, pe ansamblu a *n* ani:

$$S_j = \frac{1}{n}\sum_{i=1}^{n} S_{ij}$$

unde $S_{ij} = S_t, j$ este luna sau trimestrul pentru care se calculează coeficientul sezonier, iar *i* reprezintă anii observaţi.

Conform principiului compensării variaţiilor sezoniere la nivelul anului, suma, respectiv *media coeficienţilor sezonieri*, pe an, trebuie să fie zero. În calcule apar rezultate uşor diferite, ca urmare a aproximărilor. Efectul lor poate fi compensat printr-un *corector „$d_t$"*, rezultând un *coeficient sezonier corectat*, $S_j^{'}$.

- În cazul modelului aditiv, corectarea coeficienţilor sezonieri presupune calculul diferenţelor:

$$S_j^{'} = S_j - d_t,$$

unde „$d_t$" reprezintă corectorul: $d_t = \frac{1}{4}\sum_{j=1}^{4} S_j$, sau $d_t = \frac{1}{12}\sum_{j=1}^{12} S_j$, în general, $d_t = \frac{1}{p}\sum_{j=1}^{p} S_j$

Rolul coeficientului corector „*d*" este de a repartiza eroarea de aproximare pe ansamblul perioadelor, astfel devenind posibilă respectarea *principiului compensării*:

$$\sum_j S_j' = 0 \text{ sau } \overline{S'} = 0, \text{ unde } \overline{S'} = \frac{1}{p}\sum_{j=1}^{p} S_j'$$

(suma şi media coeficienţilor sezonieri sunt nule pe an, în cazul modelului aditiv).

- În cazul modelului multiplicativ, corectarea coeficienţilor sezonieri presupune calculul raportului:

$$S_j' = S_j / d_t$$

iar media lor este egală cu unitatea: $\overline{S} = 1, \quad S = \frac{1}{p}\sum_{j=1}^{p} S_j'$.

### Exemplu

Datele înregistrate trimestrial, pe durata a doi ani, cu privire la cifra de afaceri a unei firme sunt prezentate în tabelul 4. Admitem ipoteza continuităţii trendului, a stabilităţii sezoniere şi a lipsei influenţelor accidentale.

Se cere:

1. să se determine tendinţa seriei
2. să se calculeze indicii de sezonalitate şi coeficienţii sezonieri
3. să se desezonalizeze seria;
4. să se extrapoleze seria pentru trimestrul I al anului următor celor observaţi.

**Tabelul nr. 4.** *(Date convenţionale)*

| Trim. \ Anul | 1 | 2 |
|---|---|---|
| I | 1 | 2 |
| 2 | 3 | 5 |
| 3 | 4 | 7 |
| 4 | 2 | 4 |

### 1. Determinarea tendinţei

Reprezentarea grafică a seriei din tabelul 4, vezi fig. 4, evidenţiază clar o evoluţie sezonieră, cu valori maxime în trimestrul 3 şi minime în trimestrul 1. De asemenea, se observă o evoluţie medie liniară $f_t = a + bt$.

Elementele de calcul pentru linia de trend şi valorile estimate ( $y_{t_i}$ ) sunt prezentate în tabelul 5.

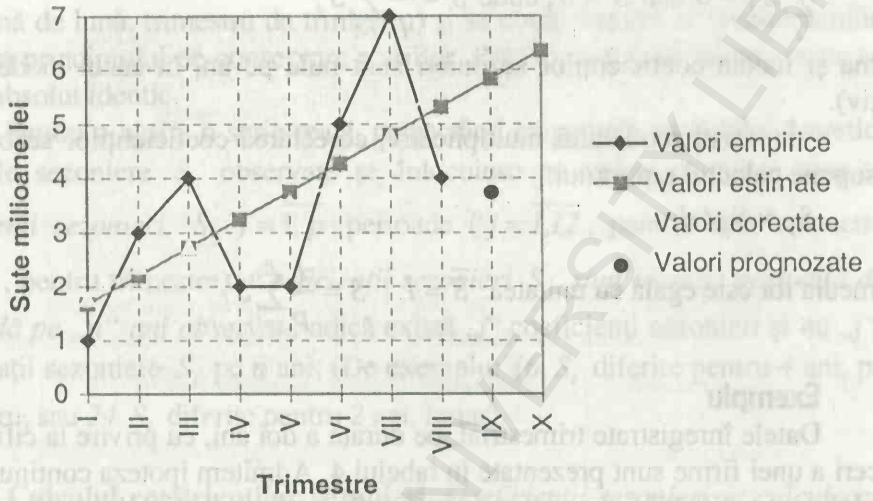



**Figura nr. 4.** *Dinamica cifrei de afaceri trimestriale (sute mil. lei) a firmei A în anii 1-2 şi prognoză pentru trimestrul 1, anul 3*

**Tabelul nr. 5.** *Elementele de calcul, valori estimate* ( $y_{t_i}$ )

| Trim.*) ($t_i$) | Valori empirice ($y_i$) | $t_i y_i$ | $t_i^2$ | Valori estimate ($y_{t_i}$) | Indici sezonieri ($i_t$) | Valori corectate |
|---|---|---|---|---|---|---|
| 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| 1 | 1 | 1 | 1 | 1,68 | 0,595 | 1,773 |
| 2 | 3 | 6 | 4 | 2,20 | 1,364 | 2,381 |
| 3 | 4 | 12 | 9 | 2,72 | 1,470 | 2,732 |
| 4 | 2 | 8 | 16 | 3,24 | 0,617 | 2,920 |
| 5 | 2 | 10 | 25 | 3,76 | 0,532 | 3,546 |
| 6 | 5 | 30 | 36 | 4,28 | 1,168 | 3,949 |
| 7 | 7 | 49 | 49 | 4,80 | 1,458 | 4,781 |
| 8 | 4 | 32 | 64 | 5,32 | 0,752 | 5,839 |
| $\sum t_i$ 36 | $\sum y_i$ 28 | $\sum t_i y_i$ 148 | $\sum t_i^2$ 204 | $\sum y_{t_i}$ 28 | - | 27,92 |

*) *Nu se mai anulează influenţa factorului timp*

BCU IASI / CENTRAL UNIVERSITY LIBRARY

Calculul parametrilor:

$$b = \frac{n\sum t_i y_i - \sum t_i \sum y_i}{n\sum t_i^2 - (\sum t_i)^2} = \frac{8 \cdot 148 - 36 \cdot 28}{8 \cdot 204 - (28)^2} = 0{,}52$$

$$a = \frac{\sum y_i}{n} - \hat{b}\frac{\sum t_i}{n} = \frac{28}{8} - 0{,}52\frac{36}{8} = 1{,}16$$

Ecuația de trend liniar pentru seria considerată este: $y_t = 1{,}16 + 0{,}52t$. Valorile calculate, $y_{ti}$, sunt prezentate în tabelul 6, coloana 5. Se verifică condiția de normalitate a trendului: $\sum y_i = \sum y_{t_i}$.

## 2. Calculul indicilor de sezonalitate și a coeficienților de sezonalitate

**Indicii de sezonalitate** sunt calculați ca raport între valoarea observată $y_i$ și valoarea calculată corespunzătoare a trendului $f_t$, respectiv, $y_{ti}$. Rezultatele sunt prezentate în tabelul 6, coloana 6.

Se observă că indicii de sezonalitate variază în jurul unității. Valoarea lor medie la nivelul unui ciclu sezonier (al unui an, în cazul nostru) este egală cu unitatea. De asemenea, se observă că valorile primului și ultimului trimestru, din fiecare an, sunt subunitare, în celelalte trimestre sunt supraunitare, și că valorile lor din fiecare trimestru sunt diferite.

Coeficienții de sezonalitate se calculează ca medie aritmetică simplă a variațiilor sezoniere pentru fiecare trimestru ($S_j$), în cursul celor doi ani considerați (cicluri sezoniere) și anume:

$$S_1 = \frac{0{,}595 + 0{,}532}{2} = 0{,}564$$

$$S_2 = \frac{1{,}364 + 1{,}168}{2} = 1{,}266$$

$$S_3 = \frac{1{,}470 + 1{,}458}{2} = 1{,}464$$

$$S_4 = \frac{0{,}617 + 0{,}752}{2} = 0{,}685$$

*Observații*

Media celor patru coeficienți de sezonalitate trebuie să fie egală cu 1, evidențiind faptul că variațiile sezoniere în interiorul unui ciclu se compensează. În cazul nostru, valoarea medie a coeficienților de sezonalitate este egală cu 0,995, valoare ce poate fi admisă, ținând cont de aproximările luate în calcul.

**3. Desezonalizarea seriei**

Desezonalizarea seriei presupune calculul valorilor corectate şi are ca scop obţinerea tendinţei fără influenţa sezonieră. Seria desezonalizată se obţine prin raportarea valorilor empirice, $y_i$, la valoarea coeficienţilor de sezonalitate corespunzători, ($S_j$). Rezultatele sunt prezentate în tabelul 5, coloana 7 şi în figura 4.

**4. Prognoza nivelului cifrei de afaceri pentru trimestrul 1 al anului următor celor observaţi.**

Folosim modelul de compunere multiplicativ, $y_t = f_t \cdot S_t$,

unde: $f_t$- trendul; $S_t$ - componenta sezonieră.

**a.** Extrapolarea trendului. Valoarea prognozată, prin extrapolarea trendului, pentru $t_i = 9$ (trimestrul 1 al anului 3), este:

$y_9 = 1{,}16 + 0{,}52 \cdot 9 = 5{,}84$.

**b.** Corectarea valorii prognozate. Corectarea valorii extrapolate cu influenţa sezonieră presupune *resezonalizarea valorii*, adică:

$$y_9^{(1)} = y_9 \cdot \bar{I}_1 = 5{,}84 \cdot 0{,}564 = 3{,}7376$$

În concluzie, ne putem aştepta ca cifra de afaceri a firmei „A" să atingă, în trimestrul 1 al anului trei considerat, valoarea de 3,7376 sute milioane lei, numai dacă se respectă ipotezele admise iniţial, şi anume: continuitatea trendului, păstrarea stabilităţii sezoniere şi lipsa influenţelor accidentale.

## 6.4. Metode moderne de analiză a seriilor de timp

Metodele moderne de analiză tratează seria de timp ca o realizare a unui proces stochastic. Distincţia dintre un proces stochastic şi o realizare a acestuia se poate face în mod analog cu distincţia dintre populaţie şi eşantion din statistica inferenţială.

Procesul stochastic sau procesul aleator reprezintă instrumentul de analiză a unui fenomen care evoluează în timp sub influenţa acţiunii unor factori aleatori, sintetizaţi într-o variabilă reziduală.

Variabila aleatoare poate fi privită fie ca o influenţă perturbatoare asupra unei legi de evoluţie a fenomenului sau poate constitui însăşi natura evoluţiei fenomenului.

Un exemplu de proces aleator este procesul gaussian numit „zgomot alb" sau proces pur aleator, exprimat prin:

$\{Y_t = \varepsilon_t, t \in T\}$, unde $\varepsilon_t$ este o variabilă aleatoare normal distribuită, $\varepsilon_t \sim N(0, \sigma^2)$, $t$ este momentul de timp.

## 6.4.1 Procese stochastice

Un proces stochastic este definit de o familie de variabile aleatoare care depind de variabila timp: $\{Y_t, t \in T\}$.

Considerăm un set de date, $(y_1, y_2, ..., y_T)$, care reprezintă o posibilă realizare a procesului stochastic.

În general, pentru $n$ posibile realizări ale procesului stochastic, se consideră un eşantion de forma $\{y_{1t}, y_{2t}, ..., y_{nt}\}$, adică un eşantion de $n$ realizări ale variabilei aleatoare $Y_t$. Variabila $Y_t$ are o densitate de probabilitate, $f_{Y_t}(y_t)$, numită densitate necondiţionată a lui $Y_t$.

De exemplu, pentru procesul gaussian „zgomot alb", $Y_t = \varepsilon_t$, putem construi funcţia densitate de probabilitate: $f_{Y_t}(y_t) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{y_t^2}{2\sigma^2}}$.

Orice proces stochastic are o repartiţie de probabilitate determinată de o familie de repartiţii finit dimensionale.

### Media şi varianţa unui proces stochastic

Valoarea medie a procesului stochastic este $\mu_t = M(Y_t)$, $t \in T$.

$M(Y_t) = \int_R y_t f_{Y_t}(y_t)\, dy_t$ (pentru un proces continuu) sau

$M(Y_t) = \lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} Y_t^{(i)}$ (pentru un proces discret).

Varianţa procesului stochastic $Y_t$ se notează cu $\gamma_{0t}$ şi este definită prin următoarea relaţie:

$V(Y_t) = \gamma_{0t} = M(Y_t - \mu_t)^2$.

### Covarianţa şi funcţia de autocorelaţie

Covarianţa sau autocovarianţa la un decalaj $k$ reprezintă covarianţa dintre variabilele $Y_t$ şi $Y_{t-k}$ :

$\gamma_{kt} = M(Y_t - \mu_t)(Y_{t-k} - \mu_{t-k})$

Evident, pentru $k = 0$, $\gamma_{kt} = \gamma_{0t} = M(Y_t - \mu_t)^2$, adică este varianţa procesului.

Funcţia de autocorelaţie la un decalaj $k$ este dată prin: $\rho_k = \frac{\gamma_{kt}}{\gamma_{0t}}$, şi poate lua valori în intervalul: $-1 \le \rho_k \le 1$.

### Exemple de procese stochastice

$Y_t = \mu + \varepsilon_t \Rightarrow M(Y_t) = \mu,\ \gamma_{0t} = \sigma^2$

$Y_t = \beta t + \varepsilon_t \Rightarrow M(Y_t) = \beta t,\ \gamma_{0t} = M(Y_t - \mu_t)^2 = M(\varepsilon_t^2) = \sigma^2,$

unde $\varepsilon_t \sim N(0, \sigma^2)$.

### Procese staţionare

Un proces stochastic se numeşte staţionar dacă media şi varianţa sa nu depind de timp, sunt constante în timp, iar covarianţa la un decalaj $k$ nu depinde decât de acel decalaj (nu depinde de timp), adică sunt îndeplinite condiţiile:

$M(Y_t) = \mu$

$\gamma_{0t} = M(Y_t - \mu_t)^2 = \sigma^2$

$M(Y_t - \mu_t)(Y_{t-k} - \mu_{t-k}) = M(Y_t - \mu)(Y_{t-k} - \mu) = \gamma_k$

### Exemplu de proces staţionar

Procesul $Y_t = \mu + \varepsilon_t$ este staţionar, adică:

$M(Y_t) = \mu,\ \gamma_{0t} = \sigma^2,$

$\gamma_k = M(\varepsilon_t \varepsilon_{t-k}) = 0$, pentru $k \neq 0$.

### Procese nestaţionare

Dacă procesul nu îndeplineşte una dintre condiţiile de definire a procesului staţionar, atunci el se numeşte nestaţionar.

### Exemplu de proces nestaţionar

Procesul „mers aleator", definit de relaţia: $Y_t = Y_{t-1} + \varepsilon_t$, este nestaţionar.

BCU STIINTA UNIVERSITY LIBRARY

$\varepsilon_t \sim N(\mu, \sigma^2)$, $M(\varepsilon_t \varepsilon_{t-k}) = 0$

$M(Y_t) = \mu_t$, iar $V(Y_t) = \sigma_t^2$.

Pentru a demonstra nestaţionaritatea procesului, considerăm seria:

$Y_0 = 0$

$Y_1 = \varepsilon_1$

$Y_2 = \varepsilon_1 + \varepsilon_2$

...

$Y_t = \sum \varepsilon_t$

Rezultă:

$M(Y_t) = M(\sum \varepsilon_t) = t \cdot \mu = \mu_t$, iar $V(Y_t) = V(\sum \varepsilon_t) = t \cdot \sigma^2$.

## Efecte ale nestaţionarităţii asupra modelării

Ipotezele clasice din econometrie se referă la procese staţionare:

- media procesului (estimată prin media empirică) este constantă, nu depinde de timp;
- varianţa procesului (estimată prin media empirică) este constantă, nu depinde de timp;
- autocovarianţele procesului (covarianţele dintre un moment prezent şi momentele de timp trecute) nu depind de timp.

În practică, o mare parte a variabilelor economice care depind de timp sunt nestaţionare. Nestaţionaritatea poate fi:

- deterministă (variabilele economice sunt funcţii de timp), de exemplu, modelul $Y_t = \mu + \beta t + \varepsilon_t$;
- stochastică (variabilele sunt procese autoregresive), de exemplu, modelul $Y_t = \mu + \rho Y_{t-1} + \varepsilon_t$.

Nestaţionaritatea unei serii de timp ridică atât probleme de inferenţă statistică (estimarea unei medii sau varianţe care este variabilă în timp), precum şi probleme de modelare econometrică (regresia falsă).

## Regresia falsă

În econometrie, pentru variabilele care depind de timp şi sunt nestaţionare (în sens determinist sau stochastic), apare o problemă deosebit de serioasă, numită de specialişti *regresie artificială sau falsă* („spurrious

regressions"). Dacă considerăm două variabile nestaționare și construim un model de regresie pe baza lor, atunci poate să apară un fenomen artificial de corelație între cele două variabile (în practică, se obțin coeficienți de corelație mari pentru astfel de situații), deși, în mod logic cele două variabile sunt independente.

### Testarea staționarității

Dacă se constată că un proces este nestaționar, acesta poate fi staționarizat în anumite condiții (de exemplu, prin operatorul diferență de ordinul întâi).

Înainte de testare este foarte important să se cunoască tipul de nestaționaritate al seriei (determinist sau stochastic).

Pentru a testa dacă o serie de timp este staționară, în analiza econometrică, se folosesc mai multe metode: **corelograma și funcția de autocorelație, testul Ljung-Box.**

### Corelograma și funcția de autocorelație

Etapele testării staționarității pot fi rezumate astfel:

**1.** Se determină funcția de autocorelație pentru un decalaj *k*:

$$\hat{\rho}_k = \frac{\hat{\gamma}_k}{\hat{\gamma}_0}, \text{ unde } \hat{\gamma}_k = \frac{\sum (Y_t - \bar{Y})(Y_{t-k} - \bar{Y})}{n}, \ \hat{\gamma}_0 = \frac{\sum (Y_t - \bar{Y})^2}{n}$$

**2.** Se reprezintă grafic corelograma funcției de autocorelație pentru un decalaj *k* la nivelul eșantionului.

**3.** Se analizează semnificația statistică a funcției de autocorelație $\hat{\rho}_k$. Pentru un proces pur aleator (zgomot alb) funcția de autocorelație are o distribuție aproximativ normală cu medie zero și varianță $1/n$.

Ipoteza nulă este $H_0$: $\rho_k = 0$ (proces staționar).

Decizia: dacă $-1{,}96\frac{1}{\sqrt{n}} \le \hat{\rho}_k \le 1{,}96\frac{1}{\sqrt{n}}$, atunci se acceptă ipoteza $H_0$ cu o probabilitate de 0,95.

În SPSS, corelograma prezintă intervalul de acceptare a ipotezei nule.

### Testul Ljung-Box

Se utilizează statistica *LB* pentru un eşantion de volum *n* şi un decalaj *m*, care urmează o lege de repartiţie hi-pătrat de *m* grade de libertate:

$$LB = n(n+2)\sum_{k=1}^{m}\frac{\hat{\rho}_k^2}{n-k} \sim \chi^2(m)$$

Ipoteza nulă $H_0$: $\rho_k = 0$ (proces staţionar) este respinsă dacă valoarea calculată a testului este mai mare decât valoarea teoretică pentru un $\alpha$ stabilit şi *m* grade de libertate.

### Proces integrat sau serie de timp integrată

Un procedeu de staţionarizare a unei serii de timp este aplicarea operatorului de diferenţă asupra seriei.

O serie de timp se numeşte *serie integrată de ordinul întâi* dacă prin aplicarea operatorului diferenţă de ordinul întâi se obţine o serie staţionară.

Dacă o serie de timp este diferenţiată până la ordinul *d*, iar în urma diferenţierii se obţine o serie staţionară, serie de timp iniţială se numeşte integrată de ordin *d*.

## 6.4.2 Tipuri de procese stochastice

### a. Procese autoregresive (AR)

Considerăm un proces stochastic $Y_t$. Dacă procesul este staţionar, atunci poate fi modelat prin:

$$Y_t = \alpha_0 + \alpha_1 Y_{t-1} + \varepsilon_t .$$

Modelul de mai sus este un model *AR(1)* sau un model autoregresiv de ordinul întâi.

Modelul poate fi scris şi în forma:

$$Y_t = \alpha_1 Y_{t-1} + \varepsilon_t ,$$

unde $\varepsilon_t \sim N(0,\sigma^2)$, iar $M(Y_t) = \mu$.

Modelul autoregresiv mai poate fi scris şi astfel:

$$Y_t - \mu = \alpha_1 (Y_{t-1} - \mu) + \varepsilon_t$$

Prin generalizare, se obţine modelul:

$$Y_t = \alpha_0 + \alpha_1 Y_{t-1} + \alpha_2 Y_{t-2} + ... + \alpha_p Y_{t-p} + \varepsilon_t \text{ - model } AR(p)$$

sau $$Y_t = \alpha_1 Y_{t-1} + \alpha_2 Y_{t-2} + ... + \alpha_p Y_{t-p} + \varepsilon_t$$

Dacă notăm cu $D$ operatorul decalaj, putem rescrie modelul:

$$(1 - \alpha_1 D - \alpha_2 D^2 - ... - \alpha_p D^p) Y_t = \varepsilon_t$$

Corelograma pentru un proces autoregresiv de ordinul $p$ descreşte exponenţial.

## b. Procese de tip medie mobilă (MA)

Procesele de tip medie mobilă sunt construite pe baza variabilei reziduale cu un anumit decalaj.

De exemplu, un model de tip medie mobilă de ordinul 1, care se simbolizează *MA(1)*, ia forma:

$$Y_t = \mu + \varepsilon_t + \beta_1 \varepsilon_{t-1}$$

Prin generalizare, se obţine un model medie mobilă de ordinul $q$:

$Y_t = \mu + \varepsilon_t + \beta_1 \varepsilon_{t-1} + ... + \beta_q \varepsilon_{t-q}$, simbolizat *MA(q)*.

Pe baza operatorului decalaj, modelul poate fi scris astfel:

$$(1 + \beta_1 D + \beta_2 D^2 + ... + \beta_q D^q) \varepsilon_t = Y_t$$

Parametrii procesului sunt:

$M(Y_t) = \mu$

$V(Y_t) = \gamma_{0t} = (1 + \beta_1^2)\sigma^2$

Dacă $\varepsilon_t \sim N(0, \sigma^2)$, atunci

$$\gamma_{0t} = M(Y_t - \mu)^2 = M(\varepsilon_t + \beta_1 \varepsilon_{t-1})^2 = (1 + \beta_1^2)\sigma^2$$

$$\gamma_{1t} = M(Y_t - \mu)(Y_{t-1} - \mu) = M(\varepsilon_t + \beta_1 \varepsilon_{t-1})(\varepsilon_{t-1} + \beta_1 \varepsilon_{t-2}) = \cdot \beta_1 \sigma^2$$

BCU IAŞI CENTRAL UNIVERSITY LIBRARY

$$\gamma_{kt} = M(Y_t - \mu)(Y_{t-k} - \mu) = M(\varepsilon_t + \beta_1\varepsilon_{t-1})(\varepsilon_{t-k} + \beta_1\varepsilon_{t-k+1}) = 0$$

$$\rho_1 = \frac{\beta_1}{1+\beta_1^2}$$

Corelograma procesului *MA(q)* are primii termeni semnificativ diferiți de zero.

Corelograma parțială (obținută pe baza coeficienților de corelație parțială) este caracterizată printr-o descreștere exponențială.

## c. Procese ARMA

Procesele *ARMA* sunt procese mixte, construite prin combinarea proceselor *AR* și *MA* și sunt simbolizate prin *ARMA (p,q),* unde:
*p* este ordinul autoregresiv;
*q* este ordinul pentru media mobilă.

Un exemplu:
$Y_t = \theta + \alpha_1 Y_{t-1} + \varepsilon_t + \beta_1\varepsilon_{t-1}$ - model *ARMA(1,1)*

Prin generalizare, se obține modelul:
$Y_t = \theta + \alpha_1 Y_{t-1} + \alpha_2 Y_{t-2} + ... + \alpha_p Y_{t-p} + \varepsilon_t + \beta_1\varepsilon_{t-1} + ... + \beta_q\varepsilon_{t-q}$, care este un proces *ARMA(p,q).*

## d. Procese ARIMA

Procesele de tip ARIMA se obțin pe baza proceselor ARMA care urmează și un proces de integrare de un anumit ordin.

Un model de tip ARIMA este simbolizat *ARIMA(p,d,q),* unde:
*p* este ordinul autoregresiv;
*d* este ordinul de integrare;
*q* este ordinul pentru media mobilă.

Modelele ARIMA sunt modelele de cea mai mare generalitate. Toate procesele prezentate anterior sunt cazuri particulare ale unui proces ARIMA:

*ARIMA(0,0,q) = MA(q)*
*ARIMA(p,0,0) = AR(p)*
*ARIMA(p,0,q) = ARMA(p,q)*

### 6.4.3 Prognoză prin metoda Box-Jenkins

Prognoza prin metoda Box-Jenkins presupune parcurgerea următoarelor etape: identificarea modelului, estimarea parametrilor modelului, testarea parametrilor şi realizarea prognozelor.

#### Identificarea modelului

Cu ajutorul funcţiilor de autocorelaţie şi autocorelaţie parţială se identifică *p, d, q,* respectiv parametrii care definesc tipul de proces aleator. În acest scop, se urmează etapele:

1. Se testează staţionaritatea. Dacă seria nu este staţionară, atunci se staţionarizează şi se obţine, prin operatorul diferenţă, o serie integrată de un anumit ordin. Ordinul operatorului de diferenţă ne dă parametrul *d.* Dacă seria este staţionară, atunci $d = 0$.

2. Se construieşte corelograma.

- Dacă corelograma descreşte exponenţial, iar corelograma parţială se diminuează lent, adică are primii *p* termeni sunt semnificativ diferiţi de zero (în afara intervalului de încredere), atunci avem un model *AR(p);*
- Dacă corelograma parţială descreşte exponenţial, iar corelograma se diminuează lent, adică are primii *q* termeni semnificativ diferiţi de zero (în afara intervalului de încredere), atunci avem un model *MA(q);*
- Dacă corelograma şi corelograma parţială descresc exponenţial, atunci avem un model *ARMA(p,q).*

#### Estimarea parametrilor modelului

După ce s-a identificat tipul de proces stochastic, pentru a realiza modelarea seriei de timp se estimează parametrii modelului propus.

#### Testarea

În această etapă se analizează semnificaţia parametrilor estimaţi şi se verifică dacă modelul este bun. O modalitate simplă de verificare constă în utilizarea intervalului de încredere pentru funcţiile de autocorelaţie şi autocorelaţie parţială a variabilei eroare estimate. Dacă valorile celor două funcţii pentru un număr suficient de lag-uri sunt în intervalul de încredere, atunci modelul este validat.

### Prognoză

Prognoza presupune aflarea unei estimații punctuale sau estimarea unui interval de încredere pentru variabila $Y$ la un moment viitor $t+k$.

## 6.5. Exemple în SPSS

### Exemplul 1

Datele înregistrate trimestrial, în perioada 2000 – 2005, cu privire la cifra de afaceri (în mii euro) a firmei Alpha, sunt prezentate în Tabelul 6.

**Tabelul nr. 6.** *(Date convenționale)*

| Trimestrul / Anul | I | II | III | IV |
|---|---|---|---|---|
| 2000 | 2,39 | 5,03 | 7,12 | 5,56 |
| 2001 | 7,00 | 6,26 | 8,75 | 8,43 |
| 2002 | 9,97 | 10,54 | 11,80 | 13,41 |
| 2003 | 12,60 | 15,20 | 19,11 | 16,96 |
| 2004 | 17,32 | 20,81 | 20,79 | 20,78 |
| 2005 | 22,92 | 23,92 | 25,09 | 24,27 |

Se cere:

- Să se determine tendința de evoluție a cifrei de afaceri;
- Să se extrapoleze seria pentru trimestrul I al anului următor celor observați.

### Rezolvare

**Determinarea tendinței** unei serii presupune următoarele etape:

1. reprezentarea grafică a seriei folosind cronograma liniară;
2. alegerea vizuală a modelului/modelelor de trend care aproximează tendința;
3. estimarea parametrilor modelului/modelelor de trend ales/alese;
4. alegerea celui mai bun model de trend;
5. testarea parametrilor modelului de trend ales;
6. testarea normalității erorilor modelului de regresie.

BCU "LUCIAN BLAGA" CENTRAL UNIVERSITY LIBRARY

**1.** Obținerea cronogramei liniare în SPSS presupune calea: meniul *Graphs* → comanda *Sequence* (Figura 5).

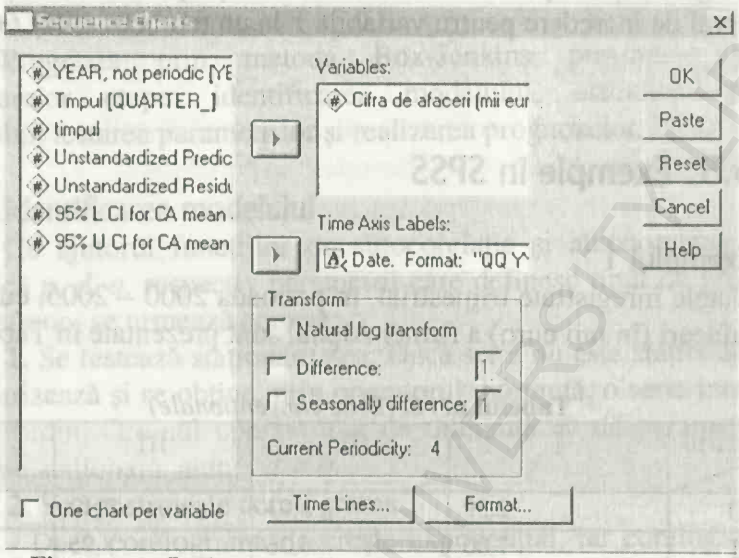



**Figura nr. 5.** *Fereastra Sequence Charts (meniul Graphs)*

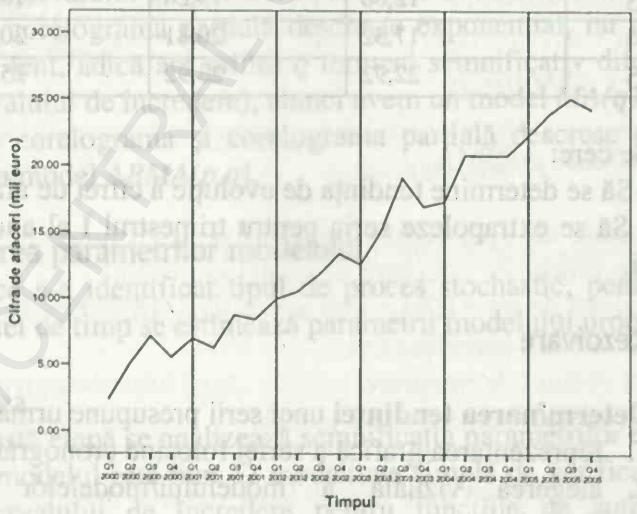



**Figura nr. 6.** *Evoluția cifrei de afaceri trimestriale a firmei Alpha în perioada 2000 – 2005*

În Output se prezintă corelograma din figura 6. Se observă o evoluție cu trend și sezonalitate.

BCU IASI CENTRAL UNIVERSITY LIBRARY

**2.** Observând corelograma, se constată o evoluție cu trend ascendent, care poate fi aproximat printr-un model de trend de tip: liniar, parabolic sau putere.

**3.** Estimarea fiecăruia din cele trei modele de trend considerate se realizează pe baza metodei celor mai mici pătrate, parcurgând următorii pași în SPSS: meniul *Analyze* → comanda *Regression* → opțiunea *Curve Estimation* (Figura 7).

Pentru variabila dependentă selectăm *cifra de afaceri,* iar pentru variabila dependentă selectăm *timpul*. Din caseta *Models* se aleg modelele: *Linear*, *Quadratic* și *Power*. De asemenea, se bifează casetele de validare *Include constant in equation* și *Plot models*. Se activează butonul OK, care comandă în SPSS obținerea output-ului dorit (Figura 8 și Tabelul 7).

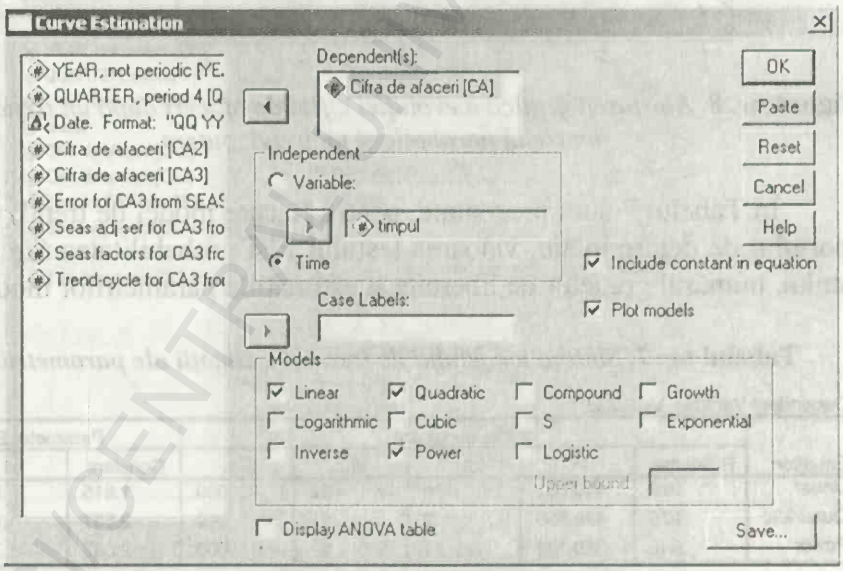



**Figura nr. 7.** *Fereastra Curve Estimation (meniul Analyze → comanda Regression)*

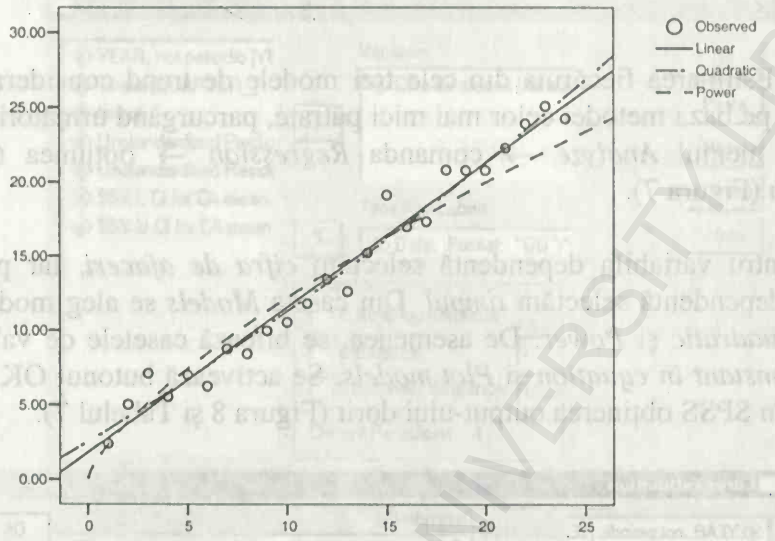



**Figura nr. 8.** *Ajustarea grafică a evoluţiei cifrei de afaceri după un trend liniar, un trend parabolic şi un trend putere*

În Tabelul 7 sunt prezentate, pentru fiecare model de trend, valoarea raportului de determinaţie, valoarea testului F şi probabilitatea *Sig* asociată testului, numărul gradelor de libertate şi estimaţiile parametrilor modelului.

**Tabelul nr. 7.** *Sinteza modelului de trend şi estimaţii ale parametrilor*

Dependent Variable: VAR00001

| | Model Summary | | | | | Parameter Estimates | | |
|---|---|---|---|---|---|---|---|---|
| Equation | R Square | F | df1 | df2 | Sig. | Constant | b1 | b2 |
| Linear | .973 | 806.651 | 1 | 22 | .000 | 1.815 | .973 | |
| Quadratic | .975 | 412.838 | 2 | 21 | .000 | 2.537 | .806 | .007 |
| Power | .942 | 360.139 | 1 | 22 | .000 | 2.357 | .713 | |

**4.** Alegerea celui mai bun model de trend. Raportul de determinaţie ($R^2$) are valori mari, apropiate de valoarea maximă 1, pentru toate modelele estimate. Modelul parabolic este modelul de trend cu valoarea cea mai mare a raportului $R^2$. Diferenţa dintre raporturile de determinaţie ale modelelor liniar şi parabolic este nesemnificativă şi potrivit **criteriului de minimizare a numărului de parametri ai modelului**, se optează pentru trendul liniar.

BCU IASI / CENTRAL UNIVERSITY LIBRARY

Estimațiile parametrilor modelului de trend liniar se calculează după relațiile:

$$b = \frac{\Delta b}{\Delta} = \frac{n\sum_i t_i \cdot y_i - \sum_i t_i \cdot \sum_i y_i}{n\sum_i t_i^2 - (\sum_i t_i)^2} = \frac{24 \cdot 5311{,}27 - 300 \cdot 335{,}4}{24 \cdot 4900 - 300^2} = 0{,}973$$

$$a = \frac{\sum_i y_i}{n} - b\frac{\sum_i t_i}{n} = \frac{335{,}4}{24} - 0{,}973\frac{300}{24} = 1{,}815$$

Ecuația estimată a modelului de trend liniar este: $y_t = 1{,}815 + 0{,}973 \cdot t$.

**5.** Testarea parametrilor modelului de trend liniar cu SPSS presupune **parcurgerea următorilor pași**: meniul *Analyze* → comanda *Regression* → opțiunea *Linear*. Fereastra *Linear Regression* este prezentată în Figura 9.

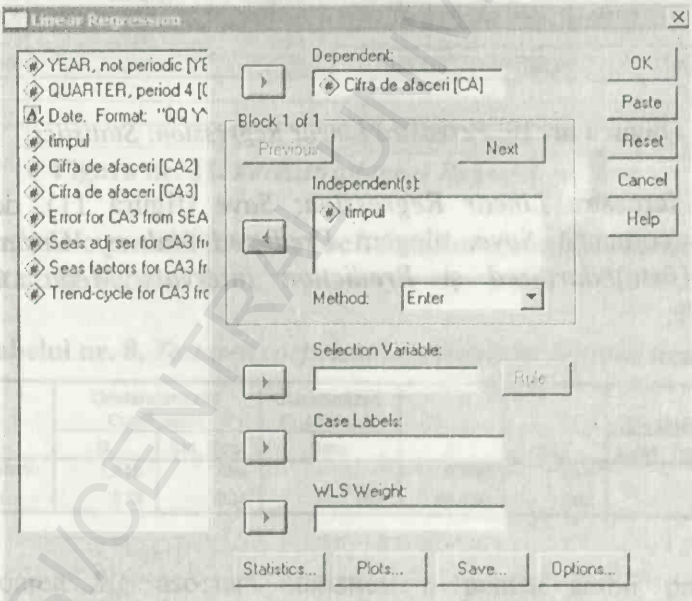



**Figura nr. 9.** *Fereastra Linear Regression (meniul Analyze → comanda Regression)*

Considerăm ca variabilă dependentă variabila *cifra de afaceri*, iar ca variabilă independentă, variabila *timp*. Prin butonul de comandă *Statistics* deschidem fereastra *Linear Regression: Statistics* (Figura 10), din care

BCU IAȘI CENTRALA UNIVERSITARĂ LIBRARY

bifăm în casetele de validare *Regression Coefficients: Estimates, Confidence intervals* şi *Model Fit.*

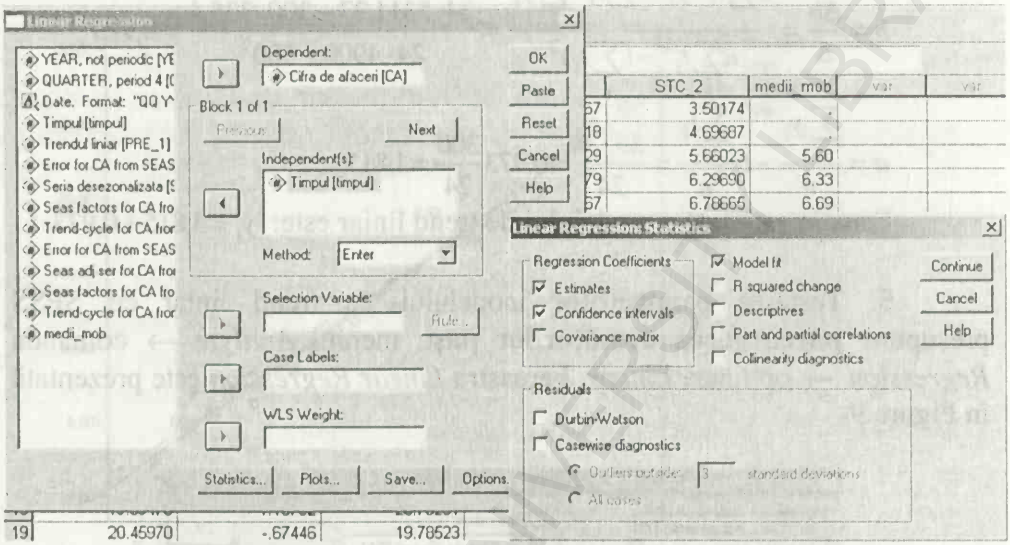



**Figura nr. 10.** *Fereastra Linear Regression: Statistics*

Din fereastra *Linear Regression: Save* (figura 11), deschisă cu butonul de comandă *Save*, alegem *Predicted Values: Unstandardized, Residual: Unstandardized* şi *Prediction Intervals Mean: Confidence Interval* 95%.

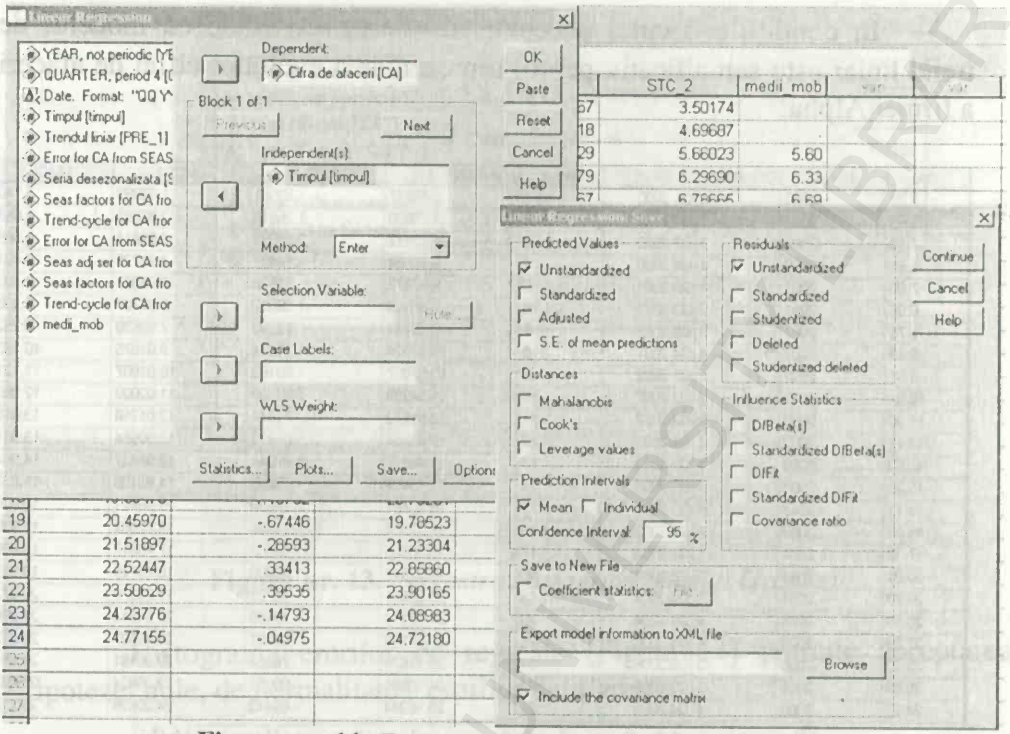



**Figura nr. 11.** *Fereastra Linear Regression: Save*

Rezultatele privind testarea coeficienților modelului de trend liniar sunt prezentate în Tabelul 8.

**Tabelul nr. 8.** *Testarea coeficienților modelului de trend liniar*

| Model | | Unstandardized Coefficients | | Standardized Coefficients | t | Sig. | 95% Confidence Interval for B | |
|---|---|---|---|---|---|---|---|---|
| | | B | Std. Error | Beta | | | Lower Bound | Upper Bound |
| 1 | (Constant) | 1.815 | .489 | | 3.708 | .001 | .800 | 2.830 |
| | timpul | .973 | .034 | .987 | 28.402 | .000 | .902 | 1.044 |

Valoarea *Sig* asociată statisticii *t*, pentru ambii parametri ai modelului de regresie, este mai mică decât riscul $\alpha$ ales, în general $\alpha = 0{,}05$. Valorile *Sig* asociate testului pentru cei doi coeficienți ai modelului de trend sunt:

$(Sig = 0{,}001) < (\alpha = 0{,}05)$ și $(Sig = 0{,}000) < (\alpha = 0{,}05)$.

Intervalul de încredere estimat nu acoperă valoarea 0, (0,800; 2,83) pentru constantă, respectiv (0,902; 1,044) pentru parametrul $\beta$.

În condițiile riscului acceptat, $\alpha = 0{,}05$, se decide că modelul de trend liniar este semnificativ pentru reprezentarea evoluției cifrei de afaceri a firmei Alpha.

| CA | YEAR_ | QUARTER_ | DATE_ | timpul | PRE_1 | RES_1 | LMCI_1 | UMCI_1 |
|---|---|---|---|---|---|---|---|---|
| 2.39 | 2000 | 1 | Q1 2000 | 1 | 2.78747 | -.39561 | 1.83399 | 3.74095 |
| 5.03 | 2000 | 2 | Q2 2000 | 2 | 3.76030 | 1.26819 | 2.86693 | 4.65367 |
| 7.12 | 2000 | 3 | Q3 2000 | 3 | 4.73312 | 2.39093 | 3.89814 | 5.56810 |
| 5.56 | 2000 | 4 | Q4 2000 | 4 | 5.70594 | -.14427 | 4.92725 | 6.48464 |
| 7.00 | 2001 | 1 | Q1 2001 | 5 | 6.67877 | .31926 | 5.95377 | 7.40377 |
| 6.26 | 2001 | 2 | Q2 2001 | 6 | 7.65159 | -1.39092 | 6.97707 | 8.32612 |
| 8.75 | 2001 | 3 | Q3 2001 | 7 | 8.62442 | .12290 | 7.99638 | 9.25245 |
| 8.43 | 2001 | 4 | Q4 2001 | 8 | 9.59724 | -1.16834 | 9.01075 | 10.18373 |
| 9.97 | 2002 | 1 | Q1 2002 | 9 | 10.57007 | -.60163 | 10.01907 | 11.12107 |
| 10.54 | 2002 | 2 | Q2 2002 | 10 | 11.54289 | -1.00789 | 11.02009 | 12.06570 |
| 11.66 | 2002 | 3 | Q3 2002 | 11 | 12.51572 | -.71647 | 12.01258 | 13.01885 |
| 13.41 | 2002 | 4 | Q4 2002 | 12 | 13.48854 | -.07412 | 12.99554 | 13.98154 |
| 12.60 | 2003 | 1 | Q1 2003 | 13 | 14.46137 | -1.86244 | 13.96837 | 14.95437 |
| 15.20 | 2003 | 2 | Q2 2003 | 14 | 15.43419 | -.23495 | 14.93106 | 15.93732 |
| 19.11 | 2003 | 3 | Q3 2003 | 15 | 16.40701 | 2.70659 | 15.88421 | 16.92982 |
| 16.96 | 2003 | 4 | Q4 2003 | 16 | 17.37984 | -.41556 | 16.82884 | 17.93084 |
| 17.32 | 2004 | 1 | Q1 2004 | 17 | 18.35266 | -1.03007 | 17.76617 | 18.93915 |
| 20.81 | 2004 | 2 | Q2 2004 | 18 | 19.32549 | 1.48000 | 18.69745 | 19.95352 |
| 20.79 | 2004 | 3 | Q3 2004 | 19 | 20.29831 | .48721 | 19.62379 | 20.97284 |
| 20.78 | 2004 | 4 | Q4 2004 | 20 | 21.27114 | -.49189 | 20.54614 | 21.99614 |
| 22.29 | 2005 | 1 | Q1 2005 | 21 | 22.24396 | .04497 | 21.46527 | 23.02265 |
| 23.92 | 2005 | 2 | Q2 2005 | 22 | 23.21679 | .70804 | 22.38181 | 24.05176 |
| 25.09 | 2005 | 3 | Q3 2005 | 23 | 24.18961 | .90051 | 23.29624 | 25.08298 |
| 24.27 | 2005 | 4 | Q4 2005 | 24 | 25.16244 | -.89442 | 24.20896 | 26.11591 |

**Figura nr. 12.** *Baza de date completată cu noile variabile*

Baza de date este completată cu variabilele *PRE_1* (valorile estimate nestandardizate ale seriei), *RES_1* (erorile de regresie nestandardizate), LMCI_1 și UMCI_1 (limita inferioară, respectiv superioară a intervalului de încredere al valorilor estimate) este prezentată în Figura 12.

**5.** Pentru testarea **normalității erorilor de regresie** folosim metoda grafică (histograma și curba frecvențelor) și testul Kolmogorov – Smirnov.

Obținerea histogramei **în SPSS** presupune pașii: meniul *Graphs*→ comanda *Histogram*. În fereastra *Histogram* (Figura 13), selectăm variabila *Erori nestandardizate* și o trecem în caseta *Variable*. Curba frecvențelor se adaugă la histogramă prin bifarea casetei de validare *Display normal curve*.

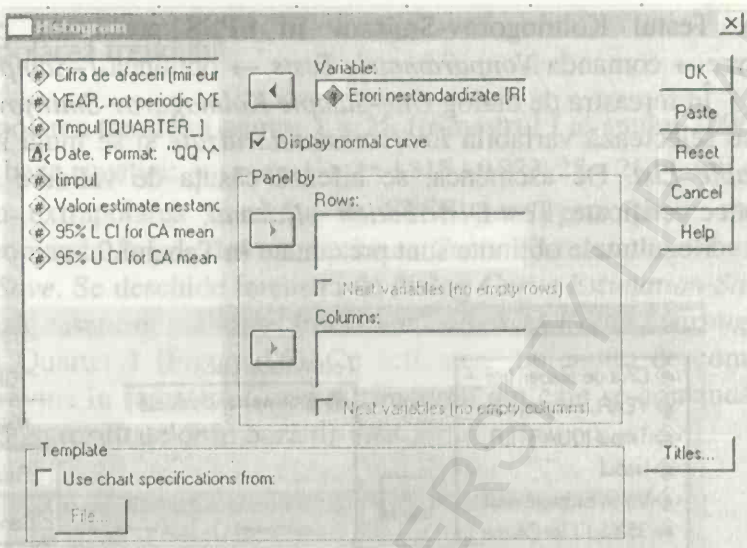



**Figura nr. 13.** *Fereastra Histogram (meniul Graphs)*

Histograma erorilor de regresie (Figura 14) permite acceptarea ipotezei nule, de normalitate a erorilor de regresie.

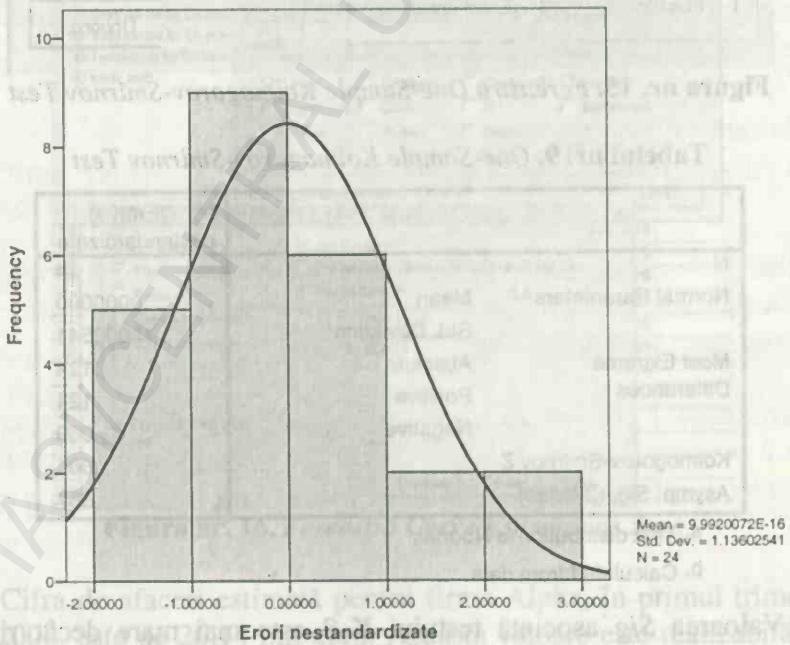



**Figura nr. 14.** *Distribuția erorilor*

BCU IAȘI CENTRAL UNIVERSITY LIBRARY

Testul Kolmogorov-Smirnov **în SPSS** presupune paşii: meniul *Analyze* → comanda *Nonparametric Tests* → opţiunea *1–Sample K-S.*

În fereastra de dialog *One-Sample Kolmogorov-Smirnov Test* (Figura 15) se selectează variabila *Erori nestandardizate* şi se mută în caseta *Test Variable List.* De asemenea, se bifează căsuţa de validare a distribuţiei teoretice verificate: *Test Distribution – Normal.*

Rezultatele obţinute sunt prezentate în Tabelul 9.

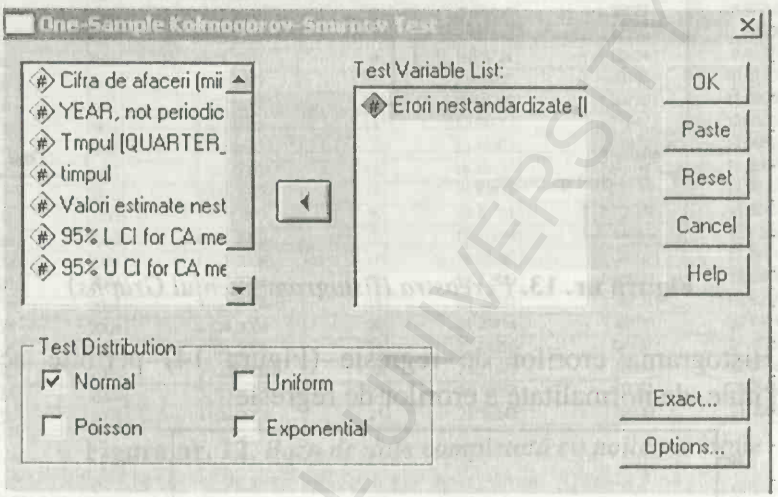



**Figura nr. 15.** *Fereastra One-Sample Kolmogorov-Smirnov Test*

**Tabelul nr. 9.** *One-Sample Kolmogorov-Smirnov Test*

| | | Erori nestandardizate |
|---|---|---|
| N | | 24 |
| Normal Parameters[a,b] | Mean | .0000000 |
| | Std. Deviation | 1.13602541 |
| Most Extreme Differences | Absolute | .124 |
| | Positive | .124 |
| | Negative | -.069 |
| Kolmogorov-Smirnov Z | | .605 |
| Asymp. Sig. (2-tailed) | | .857 |

a. Test distribution is Normal.

b. Calculated from data.

Valoarea *Sig* asociată testului K-S este mai mare decât riscul $\alpha$ acceptat, $(Sig = 0{,}857) > (\alpha = 0{,}05)$, deci se acceptă ipoteza nulă $H_0$, adică ipoteza de normalitate a erorilor.

## Extrapolarea trendului

Extrapolarea trendului pentru $t_i = 25$ (trimestrul I al anului 2006), se realizează pe baza ecuației: $y_{25} = a + b \cdot t_{25} = 1{,}815 + 0{,}973 \cdot 25 = 26{,}14$ mii euro.

Pentru extrapolarea trendului **în SPSS** parcurgem pașii: meniul *Analyze* → comanda *Regression* → opțiunea *Curve Estimation* → butonul de comandă *Save*. Se deschide fereastra de dialog *Curve Estimation Save* în care se bifează caseta de validare *Predicted values* și *Predict through* cu *Year* 2006 și Quarter 1 (Figura 16). Cu activarea butonului de comandă *Continue* se revine în fereastra *Curve Estimation*, din care se comandă *OK* pentru adăugarea rezultatelor în baza de date.

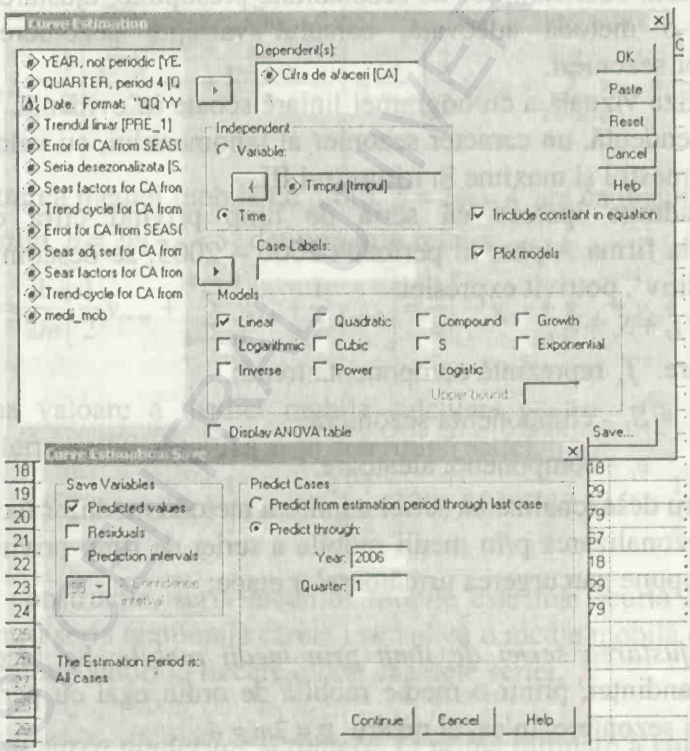



**Figura nr. 16.** *Fereastra Curve Estimations Save*

Cifra de afaceri estimată pentru firma Alpha, în primul trimestru al anului 2006, este de 26,14 mii euro. Această valoare este realizabilă numai dacă se păstrează condițiile de stabilitate a trendului și de lipsă a influențelor accidentale.

### Exemplul 2

Se consideră datele din Tabelul 6, cu privire la cifra de afaceri (în mii euro) a firmei Alpha, înregistrată trimestrial, în perioada 2000 – 2005.

Se cere:

**1.** Să se calculeze coeficienții de sezonalitate;

**2.** Să se desezonalizeze seria;

**3.** Să se extrapoleze seria pentru trimestrul I al anului următor celor observați.

### Rezolvare

#### 1. Calculul coeficienţilor de sezonalitate

Calculul coeficienților de sezonalitate presupune: ajustarea seriei de timp printr-o metodă adecvată, calculul variației sezoniere, calculul coeficienților sezonieri.

Analiza vizuală a cronogramei liniare scoate în evidență, pe lângă o tendință ascendentă, un caracter sezonier al fenomenului, cu valori minime în primul trimestru și maxime în trimestrul III.

Se admite ipoteza că seria de timp privind cifra de afaceri înregistrată la firma Alpha, în perioada 2000 – 2005, se descompune după un model aditiv[1], potrivit expresiei:

$$y_t = f_t + S_t + e_t,$$

în care: $f_t$ reprezintă componenta trend;

$S_t$ - componenta sezonieră;

$e_t$ - componenta aleatoare.

Pentru desezonalizarea seriei admitem metoda mediilor mobile.

Desezonalizarea prin medii mobile a seriei de timp privind cifra de afaceri presupune parcurgerea următoarelor etape:

*a) Ajustarea seriei de timp prin medii mobile.* Se face o primă estimare a tendinței, printr-o medie mobilă de ordin egal cu periodicitatea componentei sezoniere, în cazul nostru $p = 2m = 4$:

$$\bar{y}_{t,2m} = \frac{\sum_{k=-m}^{m-1} y_{t+k}}{2m} = \frac{y_{t-2} + y_{t-1} + y_t + y_{t+1}}{4}.$$

[1] Alegerea modelului de agregare a componentelor unei serii de timp se face pe baza testelor: testul benzii, testul Buys-Ballot.

BCU IAȘI CENTRAL UNIVERSITY LIBRARY

Mediile mobile au proprietatea de a elimina oscilațiile sezoniere de perioadă egală cu ordinul mediei mobile. Mediile mobile de ordin $p=4$ elimină oscilațiile sezoniere trimestriale, de periodicitate $P=4$; mediile mobile de ordin $p=12$ elimină sezonalitatea lunară, de periodicitate $P=12$.

Mediile mobile de ordin par sunt descentrate, asimetrice. Media mobilă nu corespunde cu termenul central al șirului pentru care s-a calculat și se produce un defazaj nedorit între dinamica seriei empirice și dinamica seriei ajustate.

Centrarea mediei mobile de ordin par, $p=2m$, constă în aplicarea la seria mediilor mobile de ordin $p$, a unei medii mobile de ordin 2. Se obține o medie mobilă centrată de ordin impar, $p'=2m+2-1=2m+1$. Cele două operații sunt sintetizate în relația:

$$\bar{y}_{t,2m+1}=\frac{1}{2}\left[\frac{\sum_{k=-m}^{m-1}y_{t+k}}{2m}+\frac{\sum_{k=-m+1}^{m}y_{t+k}}{2m}\right]=\frac{1}{2m}\left[\frac{1}{2}y_{t-m}+\sum_{k=-m+1}^{k=m-1}y_{t+k}+\frac{1}{2}y_{t+m}\right]$$

Centrarea mediei mobile de ordin $p=2m=4$ are ca rezultat o medie mobilă de ordin $p=2m+1=5$, definită prin relația:

$$\bar{y}_{t,p}=\frac{1}{2m}\left[\frac{1}{2}y_{t-m}+\sum_{k=-m+1}^{k=m-1}y_{t+k}+\frac{1}{2}y_{t+m}\right]=\frac{1}{4}\left[\frac{1}{2}y_{t-2}+y_{t-1}+y_t+y_{t+1}+\frac{1}{2}y_{t+2}\right]$$

Prima valoare a mediei mobile calculate pentru cifra de afaceri corespunde termenului al treilea și se determină astfel:

$$\bar{y}_{3,5}=\frac{1}{4}\left[\frac{1}{2}y_{t-2}+y_{t-1}+y_t+y_{t+1}+\frac{1}{2}y_{t+2}\right]=\frac{1}{4}\left[\frac{1}{2}\cdot 2{,}39+5{,}03+7{,}12+5{,}56+\frac{1}{2}7{,}00\right]=5{,}601$$

mii euro.

Prin construcție, seria mediilor mobile este mai scurtă decât seria inițială. Pentru seria temporală căreia i se aplică o medie mobilă de ordin 5, se pierd câte două valori la fiecare dintre capetele seriei.

*b) Estimarea abaterilor sezoniere.* O primă estimare a **componentei sezoniere** se face prin excluderea din seria inițială a componentei trend, estimată în prima etapă. Procedând astfel, se obține o componentă care cumulează variațiile sezoniere și aleatoare. Relația de calcul este următoarea:

$$d_i=y_i-\bar{y}_i.$$

Valoarea componentei sezoniere la momentul $t=3$ se determină astfel:

$d_3 = y_3 - \bar{y}_3 = 7{,}124 - 5{,}602 = 1{,}521$ mii euro.

*c) Calculul coeficienților sezonieri.* Coeficienții sezonieri ($S_j$) se calculează ca o medie a valorilor componentei sezoniere ($d_t$) pentru fiecare perioadă sezonieră (lună, trimestru etc.) în cursul celor *T-1* ani cuprinși în seria mediei mobile, potrivit expresiei:

$$S_j = \frac{1}{T-1}\sum_{k=1}^{T-1} d_{kj},$$

unde $d_{kj} = d_t$, $j$ reprezintă perioada sezonieră pentru care se calculează coeficientul de sezonalitate, $j = \overline{1,P}$, iar $k$ este anul, $k = \overline{1,T}$.

Coeficientul sezonier corespunzător trimestrului III se calculează astfel:

$$S_3 = \frac{1}{6-1}\sum_{k=1}^{6-1} d_{k3} = \frac{1{,}522 + 0{,}767 + 0{,}412 + 2{,}554 + 0{,}242}{5} = 1{,}000 \text{ mii euro}$$

Calculele intermediare și rezultatele sunt prezentate în output-ul obținut cu programul SPSS (Tabelul 10).

**Pași de urmat în SPSS** pentru determinarea coeficienților de sezonalitate: meniul *Analyze* → comanda *Time Series* → opțiunea *Seasonal Decomposition.*

Se deschide fereastra de dialog *Seasonal Descomposition* (Figura 17), în care se bifează casetele de validare *Additive*, corespunzător modelului de agregare considerat, și *Display casewise listing,* pentru listarea în output a tuturor variabilelor obținute în etapele intermediare (medii mobile, valorile estimate ale abaterilor sezoniere, coeficienții sezonieri finali).

Media mobilă se poate calcula în varianta necentrată, cu bifarea casetei *Moving Average Weight: All points equal*, respectiv în varianta centrată, așa cum s-a procedat în rezolvarea problemei, cu bifarea casetei *Moving Average Weight: Endpoints weghted by 0,5.*

Baza de date se completează cu variabilele nou create, procedând astfel: se activează butonul de comandă *Save* și se deschide fereastra de dialog *Season Save*, în care se bifează caseta *Add to file.*

Rezultatele obținute sunt prezentate în Tabelul 10.

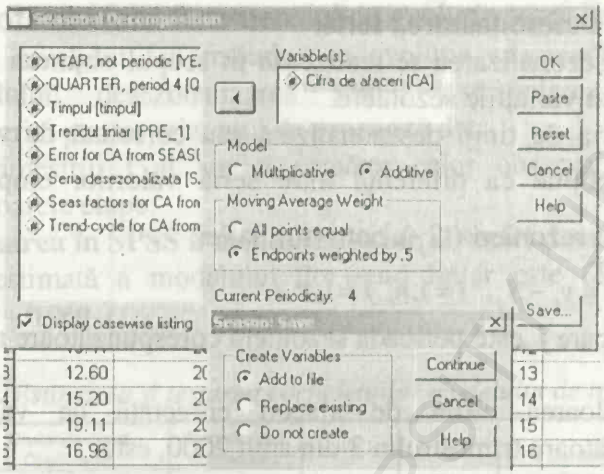



**Figura nr. 17.** Fereastra de dialog *Seasonal Decomposition*

**Tabelul nr. 10.** *Coeficienții sezonieri, seria desezonalizată, componenta reziduală*

Series Name: Cifra de afaceri

| Case Number | Original Series | Moving Average Series | Difference of Original Series from Moving Average Series | Seasonal Factor | Seasonally Adjusted Series | Smoothed Trend-Cycle Series | Irregular (Error) Component |
|---|---|---|---|---|---|---|---|
| 1 | 2.392 | . | . | -.56967 | 2.962 | 3.502 | -.54021 |
| 2 | 5.028 | . | . | .02318 | 5.005 | 4.697 | .30844 |
| 3 | 7.124 | 5.6023 | 1.52176 | 1.00029 | 6.124 | 5.660 | .46353 |
| 4 | 5.562 | 6.3321 | -.77041 | -.45379 | 6.015 | 6.297 | -.28144 |
| 5 | 6.998 | 6.6890 | .30901 | -.56967 | 7.568 | 6.787 | .78106 |
| 6 | 6.261 | 7.2503 | -.98965 | .02318 | 6.237 | 7.138 | -.90030 |
| 7 | 8.747 | 7.9800 | .76728 | 1.00029 | 7.747 | 7.954 | -.20711 |
| 8 | 8.429 | 8.8856 | -.45672 | -.45379 | 8.883 | 8.885 | -.00260 |
| 9 | 9.968 | 9.8014 | .16703 | -.56967 | 10.538 | 9.883 | .65485 |
| 10 | 10.535 | 10.8061 | -.27108 | .02318 | 10.512 | 10.773 | -.26157 |
| 11 | 11.799 | 11.7581 | .04115 | 1.00029 | 10.799 | 11.652 | -.85256 |
| 12 | 13.414 | 12.6699 | .74450 | -.45379 | 13.868 | 12.803 | 1.06514 |
| 13 | 12.599 | 14.1673 | -1.56833 | -.56967 | 13.169 | 14.056 | -.88770 |
| 14 | 15.199 | 15.5253 | -.32604 | .02318 | 15.176 | 15.486 | -.31041 |
| 15 | 19.114 | 16.5595 | 2.55413 | 1.00029 | 18.113 | 16.732 | 1.38120 |
| 16 | 16.964 | 17.8507 | -.88643 | -.45379 | 17.418 | 17.803 | -.38457 |
| 17 | 17.323 | 18.7605 | -1.43789 | -.56967 | 17.892 | 18.664 | -.77174 |
| 18 | 20.805 | 19.4463 | 1.35914 | .02318 | 20.782 | 19.595 | 1.18752 |
| 19 | 20.786 | 20.5440 | .24152 | 1.00029 | 19.785 | 20.460 | -.67446 |
| 20 | 20.779 | 21.5547 | -.77547 | -.45379 | 21.233 | 21.519 | -.28593 |
| 21 | 22.289 | 22.4827 | -.19378 | -.56967 | 22.859 | 22.524 | .33413 |
| 22 | 23.925 | 23.4569 | .46795 | .02318 | 23.902 | 23.506 | .39535 |
| 23 | 25.090 | . | . | 1.00029 | 24.090 | 24.238 | -.14793 |
| 24 | 24.268 | . | . | -.45379 | 24.722 | 24.772 | -.04975 |

## 2. Desezonalizarea seriei

Desezonalizarea se realizează în scopul obținerii unei serii de timp corectată cu variațiile sezoniere.

Seria de timp desezonalizată sau corectată de variațiile sezoniere ($y_t^*$) se obține ca diferență între seria valorilor empirice ($y_i = y_{kj}$) și coeficienții sezonieri ($S_j$), potrivit relației:

$$y_i^* = y_i - S_j, \quad i = \overline{1,n}, \quad j = \overline{1,P},$$

în care $j$ este perioada sezonieră corespunzătoare momentului $i$.

Valoarea cifrei de afaceri corectată de variațiile sezoniere, corespunzătoare trimestrului 3 din anul 2000, este:

$$y_3^* = y_3 - S_3 = 7{,}124 - 1{,}000 = 6{,}124 \text{ mii euro.}$$

Valorile pentru ceilalți termeni sunt prezentate în tabelul 10 și figura 18.

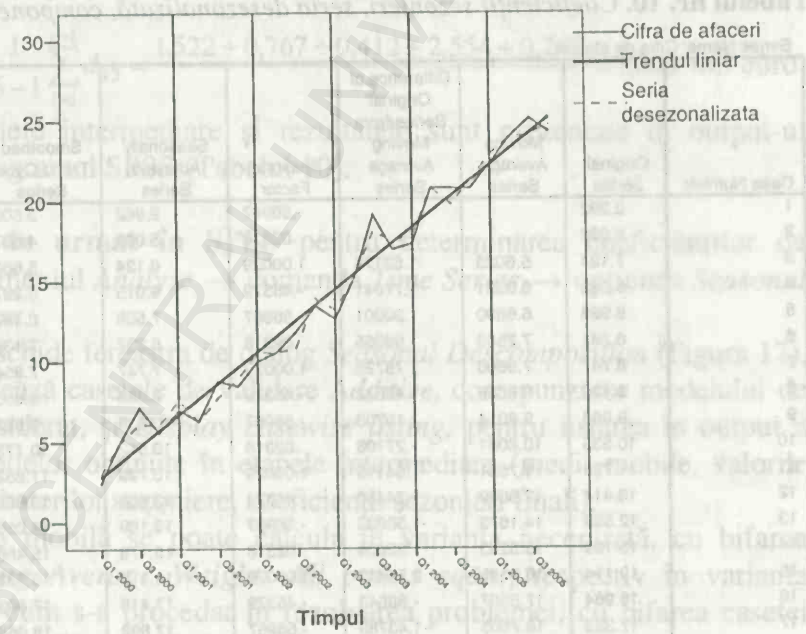



**Figura nr. 18.** *Dinamica cifrei de afaceri a firmei Alpha, în perioada 2000-2005 (Seria empirică, Seria desezonalizată, Trendul liniar)*

BCU IASI CENTRAL UNIVERSITY LIBRARY

### 3. Prognoza nivelului fenomenului cu evoluţie sezonieră

Prognoza nivelului fenomenului cu evoluţie sezonieră presupune estimarea trendului, desezonalizarea seriei, extrapolarea trendului, resezonalizarea seriei pentru orizontul de prognoză dat.

Pentru trimestrul I al anului următor celor observaţi, prognoza presupune următoarele etape:

**a) Estimarea în SPSS a trendului liniar al seriei desezonalizate**

Ecuaţia estimată a modelului de trend liniar este (Tabelul 11): $y_t = 1{,}858 + 0{,}969 \cdot t$.

**Tabelul nr. 11.** *Estimarea şi testarea coeficienţilor modelului de trend liniar*

| Model | | Unstandardized Coefficients | | Standardized Coefficients | t | Sig. | 95% Confidence Interval for B | |
|---|---|---|---|---|---|---|---|---|
| | | B | Std. Error | Beta | | | Lower Bound | Upper Bound |
| 1 | (Constant) | 1.858 | .403 | | 4.610 | .000 | 1.022 | 2.694 |
| | Timpul | .969 | .028 | .991 | 34.372 | .000 | .911 | 1.028 |

Validarea modelului de trend liniar estimat s-a realizat pe baza rezultatelor obţinute pentru testarea parametrilor de regresie (Tabelul 11) şi pentru verificarea normalităţii erorilor de regresie folosind histograma (Figura 19) şi testul Kolmogorov-Smirnov (Tabelul 12).

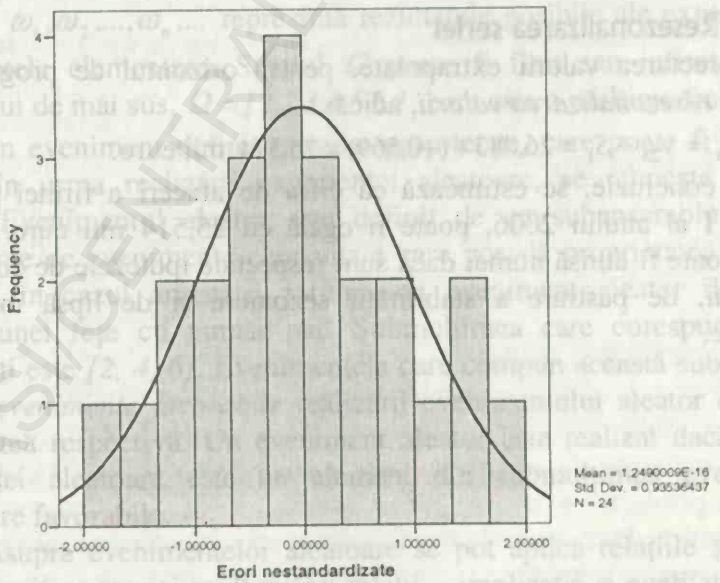



**Figura nr. 19.** *Distribuţia erorilor*

**Tabelul nr. 12.** *One-Sample Kolmogorov-Smirnov Test*

| | | Unstandardized Residual |
|---|---|---|
| N | | 24 |
| Normal Parameters[a,b] | Mean | .0000000 |
| | Std. Deviation | .93536437 |
| Most Extreme Differences | Absolute | .069 |
| | Positive | .068 |
| | Negative | -.069 |
| Kolmogorov-Smirnov Z | | .336 |
| Asymp. Sig. (2-tailed) | | 1.000 |

a. Test distribution is Normal.

b. Calculated from data.

## b. Extrapolarea trendului

Valoarea prognozată prin extrapolarea trendului, pentru $t_l = 25$ (trimestrul I al anului 2006), este:

$$y_{25} = 1{,}858 + 0{,}969 \cdot 25 = 26{,}083 \text{ mii euro.}$$

În SPSS, pentru testarea parametrilor de regresie, testarea normalității erorilor şi extrapolarea trendului, se procedează în mod analog demersului din problema 1.

## c. Resezonalizarea seriei

Corectarea valorii extrapolate, pentru orizontul de prognoză ales presupune *resezonalizarea valorii,* adică:

$$y_{25}^{(1)} = y_{25} + S_1 = 26{,}083 + (-0{,}569) = 25{,}514 \text{ mii euro.}$$

În concluzie, se estimează că cifra de afaceri a firmei Alpha, în trimestrul I al anului 2006, poate fi egală cu 25,514 mii euro. Valoarea estimată poate fi atinsă numai dacă sunt respectate ipotezele de continuitate a trendului, de păstrare a stabilității sezoniere şi de lipsă influențelor accidentale.

BCU IASI CENTRAL UNIVERSITY LIBRARY

# Anexă

## Noţiuni de teoria probabilităţilor

În această anexă, prezentăm un sumar de noţiuni statistice şi matematice utilizate în cercetarea econometrică. Vizăm, în mod deosebit, noţiunea de: probabilitate, variabilă aleatoare şi repartiţie de probabilitate.

### 1. Probabilitate

#### a. Experienţă aleatoare, eveniment aleator

Prin *experienţă aleatoare* înţelegem o acţiune sau un proces care conduce la un ansamblu de rezultate posibile, fiecare rezultat fiind supus aleatorului, adică nu poate fi anticipat. Un rezultat posibil al experienţei aleatoare se numeşte *eveniment elementar*, iar ansamblul evenimentelor elementare se numeşte *spaţiul evenimentelor elementare*, notat cu $\Omega$. Un exemplu clasic de experienţă aleatoare este aruncarea unui zar. Evenimentul elementar este apariţia unei feţe, în total, existând şase evenimente elementare.

Unei experienţe aleatoare i se asociază un spaţiu $\Omega$ ale cărui elemente $\omega_1, \omega_2, ..., \omega_n, ...$ reprezintă rezultatele posibile ale experienţei sau evenimentele elementare. Spaţiul $\Omega$ poate fi finit sau infinit. Conform exemplului de mai sus, $\Omega = \{1,2,3,4,5,6\}$, care este o mulţime finită.

Un eveniment definit printr-o proprietate, care poate fi îndeplinită sau nu în urma realizării experienţei aleatoare, se numeşte *eveniment aleator*. Evenimentul aleator este definit de un subansamblu sau de o submulţime de evenimente elementare care posedă proprietatea cerută. De exemplu, în cazul aruncării zarului, un eveniment aleator îl constituie apariţia unei feţe cu număr par. Submulţimea care corespunde acestei proprietăţi este $\{2, 4, 6\}$. Evenimentele care compun această submulţime se numesc *evenimente favorabile* realizării evenimentului aleator definit prin proprietatea respectivă. Un eveniment aleator este realizat dacă rezultatul experienţei aleatoare este un element din submulţimea evenimentelor elementare favorabile.

Asupra evenimentelor aleatoare se pot aplica relaţiile şi operaţiile logice specifice teoriei mulţimilor: relaţii – implicaţia şi egalitatea; operaţii – reuniune, intersecţie şi complementaritate.

Două evenimente aleatoare se numesc *incompatibile* sau *mutual exclusive* într-o experiență dacă nu pot fi realizate simultan, adică dacă nu au nici un eveniment elementar comun. De exemplu, evenimentele $A$ și $B$ sunt incompatibile dacă $A \cap B = \varnothing$.

O clasă de subansamble sau părți nevide ale lui $\Omega$, notată $\{P_j, j \in J\}$, unde $J$ este o mulțime cel mult numărabilă, se numește *partiție* sau *sistem de evenimente mutual exclusive și exhaustive*, dacă au loc condițiile:

- $P_i \cap P_j \neq \Phi, \ i \neq j, \ \ i,j \in J$
- $\bigcup_{j \in J} P_j = \Omega$ .

Dacă sistemul de evenimente definit mai sus face parte dintr-un corp de evenimente $K$ din $\Omega$, atunci acesta poartă numele de *sistem complet de evenimente* sau sistem *exhaustiv* de evenimente.

Un corp de evenimente $K$ se mai numește și *algebră*, și se definește prin următoarele proprietăți:

- $(\forall) \ A \in K \Rightarrow \ A^c \in K$
- $(\forall) \ A, B \in K \Rightarrow \ A \cup B \in K$
- $\Omega \in K$

Dubletul $\{\Omega, K\}$ se numește *câmp de evenimente*. Acesta este dat de mulțimea evenimentelor elementare $\Omega$ și de un corp de evenimente $K$ definit pe $\Omega$.

## b. Probabilitate

Definirea conceptului de *probabilitate* a suferit modificări de-a lungul timpului. Există două accepțiuni ale termenului de *probabilitate*: probabilitatea subiectivă și probabilitatea obiectivă.

### Probabilitatea subiectivă

Într-o primă formă, probabilitatea subiectivă implică o conotație psihologică, fiind o expresie a gradului de certitudine sau incertitudine cu privire la valoarea de adevăr a unor propoziții. Mai apoi, probabilitatea subiectivă primește și o accepțiune logică, ca o expresie a gradului de încredere rațională, care are la bază anumite cunoștințe cu privire la valoarea de adevăr a unor propoziții.

### Probabilitatea obiectivă

În această accepțiune, un enunț de probabilitate este tratat ca un enunț despre frecvența relativă a anumitor evenimente în cadrul unui șir de evenimente[1]. Pentru probabilitatea obiectivă se cunosc mai multe definiții care corespund etapelor de evoluție a conceptului în sine. Prezentăm mai jos etapele acestui proces.

**i. Definiția clasică a probabilității** este cea a lui Bernoulli și Laplace. Această definiție poate fi rezumată astfel: probabilitatea ca un eveniment să se realizeze reprezintă raportul dintre numărul de evenimente elementare favorabile realizării evenimentului și numărul evenimentelor egal posibile.

Așadar, probabilitatea este un raport dintre: cardinalul mulțimii evenimentelor favorabile și cardinalul mulțimii evenimentelor egal posibile $\Omega$.

Formalizând, avem: $p = \frac{card(A)}{card(\Omega)} = \frac{m}{n}$, unde $m$ este numărul cazurilor favorabile și $n$ este numărul cazurilor posibile. Între $m$ și $n$ există relația $0 \leq m \leq n$, adică $0 \leq p \leq 1$. Valoarea $p = 0$ corespunde imposibilității realizării evenimentului sau evenimentul imposibil, iar valoarea $p = 1$ corespunde evenimentului cert sau sigur. Evenimentul sigur este dat de mulțimea evenimentelor elementare $\Omega$. Probabilitatea obținută prin această definiție a fost numită *probabilitate a priori*.

Definiția clasică a fost criticată de specialiști (de exemplu, Poincaré) pentru că presupune „echiprobabilitatea" evenimentelor care sunt implicate într-o experiență, ceea ce reprezintă o limită sau un caz particular.

**ii. Definiția probabilității bazată pe frecvență** încearcă să înlăture critica adusă definiției clasice. Noțiunea de *frecvență* este o noțiune statistică care se bazează pe observarea și înregistrarea numărului efectiv de realizări $m$ a unui eveniment dintr-un număr $n$ de experiențe realizate. Probabilitatea este definită ca un caz limită a frecvenței, atunci când numărul de experiențe tinde la infinit. Adică, prin formalizare, se poate scrie: $p = \lim_{n \to \infty} \frac{m}{n}$. Această definiție arată că, repetând experiența de un număr foarte mare de ori, frecvența de apariție a unui eveniment se stabilizează în jurul unei valori limită care reprezintă probabilitatea de

[1] K. Popper, *Logica cercetării*, Editura Științifică și Enciclopedică, București, 1981, p. 167.

apariţie a evenimentului. Probabilitatea obţinută pe baza frecvenţei se numeşte *probabilitate a posteriori.*

Această definiţie este o expresie a Legii numerelor mari (Bernoulli) care arată că frecvenţele de apariţie a unui eveniment converg către aceeaşi valoare limită când numărul de experienţe creşte. Critica adusă acestui mod de definire a probabilităţii vizează aspectul repetitiv şi de volum mare al experienţei aleatoare.

**iii.** Conform **definiţiei axiomatice,** probabilitatea este o aplicaţie, o funcţie care asociază fiecărui eveniment dintr-un câmp de evenimente un număr real care măsoară posibilitatea ca evenimentul să apară prin realizarea experienţei aleatoare. Astfel, pe câmpul de evenimente $\{\Omega, K\}$, probabilitatea reprezintă funcţia $P: K \to R$, care îndeplineşte axiomele:

1. $(\forall)\ A \in K,\ P(A) \geq 0$
2. $P(\Omega) = 1$
3. $(\forall)\ A, B \in K,\ A \cap B = \Phi$, *atunci* $P(A \cup B) = P(A) + P(B)$ sau

$(\forall)\ \{A_i, i = \overline{1,n}\} \subset K,\ A_i \cap A_j = \Phi$, *pentru* $i \neq j$, *atunci* $P(\bigcup_{i=1}^{n} A_i) = \sum_{i=1}^{n} P(A_i)$

Tripletul $\{\Omega, K, P\}$ se numeşte *câmp de probabilitate.*

Pe baza definiţiei, se pot demonstra uşor următoarele proprietăţi ale probabilităţii:

- $(\forall)\ A \in K,\ P(A^c) = 1\text{-}P(A)$
- $P(\Phi) = 0$
- $(\forall)\ A \in K,\ 0 \leq P(A) \leq 1$
- $(\forall)\ A, B \in K,\ P(A \cup B) = P(A) + P(B)\text{-}P(A \cap B)$

## Probabilitate condiţionată

Există situaţii practice în care producerea unui eveniment este condiţionată de producerea altui eveniment. Pentru aceste situaţii, s-a definit *probabilitatea condiţionată.*

Pe un câmp de probabilitate $\{\Omega, K, P\}$, considerând evenimentele $A$, $B$ şi $P(A)>0$, se defineşte probabilitatea condiţionată prin funcţia $P_A: K \to R$, astfel încât:

$$P_A(B) = \frac{P(A \cap B)}{P(A)}.$$

Probabilitatea condiționată este o probabilitate în sensul definiției axiomatice și îndeplinește proprietățile prezentate mai sus. Probabilitatea condiționată $P_A(B)$ are la bază o informație suplimentară care este oferită de realizarea evenimentului $A$.

Din relația probabilității condiționate se deduce relația probabilității intersecției a două evenimente:

$$P(A \cap B) = P(A) \cdot P_A(B) = P(B) \cdot P_B(A).$$

### Evenimente independente

În momentul în care $P_A(B) = P(B)$, adică probabilitatea realizării evenimentului $B$ nu depinde de informația obținută prin realizarea evenimentului $A$, se afirmă că evenimentul $B$ este *independent stochastic* de evenimentul $A$.

Două evenimente A, B se numesc *independente stochastic* dacă are loc relația: $P(A \cap B) = P(A) \cdot P(B)$.

## 2. Variabile aleatoare

### A. Definiție

O experiență aleatoare este descrisă prin mulțimea evenimentelor elementare $\Omega = \{\omega_1, \omega_2, ... \omega_n, ...\}$. Această mulțime poate fi finită sau infinită.

Definirea variabilei aleatoare are la bază ideea asocierii unui număr real fiecărui eveniment elementar $\omega$ printr-o operație de măsurare. Așadar, realizăm corespondența $\omega_i \rightarrow X(\omega_i) = x_i \in R$, unde $X$ este o funcție definită pe $\Omega$ cu valori în mulțimea numerelor reale: $X : \Omega \rightarrow R$. Acestei funcții i se impune condiția de măsurabilitate, deoarece fiecărui eveniment elementar i se asociază un număr real printr-o operație de măsurare. Funcțiile reale de acest gen poartă numele de *variabile aleatoare*.

Pe câmpul de probabilitate $\{\Omega, K, P\}$, variabila aleatoare este o funcție măsurabilă, $X : \Omega \rightarrow R$.

Un exemplu de variabilă aleatoare este cea asociată experienței aleatoare a aruncării pe o masă a două zaruri. În acest caz, $\Omega = \{(i, j), i, j = \overline{1,6}\}$, iar funcția care se poate asocia experienței este aceea a atribuirii unui număr real fiecărui eveniment elementar egal cu suma punctelor obținute la fiecare aruncare. Deci,

$\omega = (i,j)$, *iar* $X(\omega) = i + j$, *pentru* $i,j = \overline{1,6}$.

Pentru a prezenta câteva dintre proprietățile unei variabile aleatoare, ne vom folosi de o clasă specială de variabile aleatoare numite variabile aleatoare simple. Pe un câmp de probabilitate $\{\Omega, K, P\}$, o asemenea variabilă ia valori distincte şi are o mulțime a valorilor finită:

$$X(\Omega) = \{x_1, x_2, ...x_n\}, \text{ cu } x_i \neq x_j \in R.$$

Evenimentele $A_i = \{\omega \in \Omega,\ X(\Omega) = x_i, i = \overline{1,n}\} \in K$ formează un *sistem complet de evenimente*, adică:

$$A_i \neq A_j, i \neq j \Rightarrow A_i \cap A_j = \Phi,\ i,j = \overline{1,n} \text{ şi } \bigcup_{i=1}^{n} A_i = \Omega.$$

În aceste condiții, notând cu $p_i = P(A_i) \Rightarrow \sum_{i=1}^{n} p_i = 1$.

Deoarece $A_i = \{\omega \in \Omega,\ X(\Omega) = x_i, i = \overline{1,n}\} \in K$, se poate vorbi despre probabilitatea ca variabila aleatoare X să ia o anumită valoare, adică:

$$p_i = P(A_i) = P(X = x_i) = P\{\omega \in \Omega,\ X(\Omega) = x_i,\ i = \overline{1,n}\}.$$

Probabilitățile $p_i$ împreună cu valorile $x_i$ determină o repartiție de probabilitate pentru variabila aleatoare $X$, care se poate scrie astfel:

$$X : \begin{pmatrix} x_i \\ p_i \end{pmatrix}_{i=\overline{1,n}}.$$

## Funcţia de repartiţie

*Repartiția sau legea de probabilitate* a unei variabile aleatoare este dată prin funcția sa de probabilitate $P^{(X)}$, care poate lua diferite forme în funcție de tipul variabilei. Pe baza funcției de probabilitate a unei variabile aleatoare se determină funcția sa de repartiție. În general, *funcția de repartiție* este definită prin relația:

$$F(x) = P^{(X)}((-\infty, x)),\ (\forall)\ x \in R \text{ sau}$$

$$F(x) = P^{(X)}[X^{-1}((-\infty, x))] = P(\{\omega \in \Omega,\ X(\omega) < x\}),\ (\forall)\ x \in R.$$

Simplificat, se mai poate scrie: $F(x) = P(X < x),\ (\forall)\ x \in R$.

Funcția de repartiție are următoarele proprietăți:

- $(\forall)\ x \in R,\ 0 \leq F(x) \leq 1$;
- $(\forall)\ a,b \in R,\ a < b,\ F(a) \leq F(b)$;
- $\lim_{x \to +\infty} F(x) = 1$;

BCU IAŞI CENTRAL UNIVERSITY LIBRARY

- $\lim_{x \to -\infty} F(x) = 0$;
- $(\forall)\ x \in R,\ F(x-0) = F(x)$.

Pentru o variabilă aleatoare X şi pentru oricare două constante reale $a$, $b$, care îndeplinesc condiţia $a < b$, se poate demonstra următorul rezultat:

$$P(a \le X < b) = F(b) - F(a),\ (\forall)\ x \in R.$$

## Variabile aleatoare discrete şi continue

Variabila aleatoare simplă, folosită ca exemplu până acum, face parte din familia variabilelor aleatoare discrete. În general, o *variabilă aleatoare discretă* ia valori distincte pe o mulţime a valorilor sale $I$, care este o mulţime cel mult numărabilă. Matematic, variabila aleatoare discretă este definită prin: $X : \begin{pmatrix} x_i \\ p_i \end{pmatrix}_{i \in I}$. Pentru această variabilă, funcţia de repartiţie este $F(x) = \sum_{\{i : x_i < x\}} p_i$.

O altă familie importantă a variabilelor aleatoare este cea a căror mulţime a valorilor este un interval mărginit sau nemărginit. Aceste variabile aleatoare se numesc continue. *Variabilele aleatoare continue* sunt date prin funcţia lor de repartiţie $F(x) = \int_{-\infty}^{x} f(t) \cdot dt$, $(\forall)\ x \in R$, unde $f(t)$ se numeşte *funcţie densitate de repartiţie.*

*Funcţia densitate de repartiţie* are următoarele proprietăţi:

- $(\forall)\ x \in R,\ f(x) \ge 0$
- $\int_{-\infty}^{+\infty} f(x) \cdot dx = 1$;
- $(\forall)\ a, b \in R,\ a < b,\ \int_{a}^{b} f(x) \cdot dx = F(b) - F(a)$;
- $f(x) = F'(x)$.

În teoria şi practica sondajelor, se folosesc o serie de repartiţii probabilistice mai importante, dintre care amintim:

- repartiţii discrete: Bernoulli, uniformă, binomială, hipergeometrică, geometrică, Poisson etc.

BCU IAŞI CENTRUL UNIVERSITAR LIBRARY

- repartiții continue: normală (standard și generalizată), beta, gama, chi-pătrat, student, Snedecor-Fisher etc.

## B. Caracteristici numerice ale unei variabile aleatoare

Distribuția unei variabile aleatoare poate fi caracterizată ușor prin intermediul unor mărimi, numite *parametri* sau *caracteristici numerice*. Prezentăm mai jos câteva dintre aceste caracteristici numerice.

### Media unei variabile aleatoare

O variabilă aleatoare $X$, definită pe câmpul $\{\Omega, K, P\}$, admite medie dacă este P – integrabilă sau P – sumabilă. În aceste condiții, media se obține prin relația:

$$M(X) = \int_R x \cdot dF(x).$$

Dacă $X$ este discretă, atunci: $M(X) = \sum_{i \in I} x_i \cdot p_i$,

iar dacă $X$ este o variabilă continuă, $M(X) = \int_R x \cdot f(x) \cdot dx$.

### Dispersia sau varianța unei variabile aleatoare

Prin definiție, dispersia unei variabile aleatoare este dată de relația:

$$D^2(X) = M[X - M(X)]^2 = M(X^2) - M^2(X),$$

unde $M(X^2)$ reprezintă media variabilei $X^2 = X \cdot X$. Existența dispersiei impune condiția ca variabila să fie pătrat sumabilă în raport cu probabilitatea.

### Abaterea medie pătratică a unei variabile aleatoare

Numită și abatere standard, abaterea medie pătratică se obține prin relația: $D(X) = \sqrt{D^2(X)}$, adică reprezintă rădăcina pătrată a dispersiei.

### Momentele unei variabile aleatoare

- *momentul inițial de ordin k* este definit prin relația: $\alpha_k(X) = M(X^k),\ k \in N$;
- *momentul absolut de ordin k* este dat prin : $\beta_k(X) = M(|X|^k),\ k \in R,\ k > 0$;
- *momentul centrat de ordin k* se obține cu relația:

$$\mu_k(X) = M\{[X - M(X)]^k\},\ k \in N.$$

### Covarianţa a două variabile aleatoare

Pentru două variabile aleatoare $X$ şi $Y$ se defineşte caracteristica numerică numită *covarianţă* sau *corelaţie prin relaţia*:

$$\text{cov}(X,Y) = M\{[X - M(X)][Y - M(Y)]\} = M(X \cdot Y) - M(X) \cdot M(Y).$$

Dacă variabilele $X$ şi $Y$ sunt independente stochastic, atunci:

$$\text{cov}(X,Y) = 0\ .$$

### Coeficientul de corelaţie a două variabile aleatoare

În ipoteza că două variabile aleatoare $X$ şi $Y$ au dispersii nenule, se defineşte coeficientul de corelaţie al celor două variabile prin relaţia:

$$\rho(X,Y) = \frac{Cov(X,Y)}{D(X) \cdot D(Y)},$$

care îndeplineşte proprietatea: $-1 \le \rho(X,Y) \le 1$.

### C. Parametri. Estimatori. Estimaţii

Teoria inferenţei statistice aduce în discuţie trei termeni fundamentali: parametrul, estimatorul şi estimaţia. Între aceşti termeni, se poate prezenta relaţia:

*parametru* $\Leftrightarrow$ *estimator* $\Leftrightarrow$ *estimaţie*

Un parametru este estimat prin intermediul unui estimator, cu ajutorul unei estimaţii. Estimaţia este valoarea calculată la nivelul unui eşantion, pe baza estimatorului, pentru a aprecia parametrul cu o anumită eroare.

**Parametrul** reprezintă o valoare fixă şi necunoscută, numită şi *valoare reală sau adevărată*, a unei populaţii studiate după o anumită caracteristică.

De exemplu, dacă notăm cu $X$ caracteristica studiată, cu $P$ populaţia de referinţă, cu N volumul populaţiei şi cu $\theta$ parametrul, atunci se poate scrie: $\theta = f(x_1, x_2, ..., x_N)$, adică parametrul, în expresia sa analitică, este definit în funcţie de valorile caracteristicii la nivelul populaţiei totale.

În general, parametrii asociaţi unei variabile aleatoare sunt caracteristicile sale numerice: media, dispersia (varianţa), abaterea medie pătratică, momentul de ordin k, momentul centrat de ordin k etc.

**Estimatorul** este o statistică, adică o variabilă aleatoare care este determinată de totalitatea eşantioanelor posibile de volum dat ($n$) care se pot extrage din populaţia de referinţă. Estimatorul, în expresia sa analitică, este definit ca o funcţie a variabilelor de selecţie $X_i$, $\overline{i=1,n}$: $\hat{\theta}=T(X_1,X_2,...,X_n)$.

Ceea ce este foarte important în cazul unui estimator este repartiţia sa probabilistică, pe baza proprietăţilor căreia se realizează inferenţa statistică.

Valorile posibile ale unui estimator se numesc **estimaţii**. O estimaţie se obţine la nivelul unui eşantion extras, pe baza datelor culese, şi este o funcţie a valorilor de sondaj $x_i$, $\overline{i=1,n}$: $\hat{\hat{\theta}}=T(x_1,x_2,...,x_n)$. Estimaţia este valoarea statistică calculată la nivelul eşantionului pe baza căreia se estimează parametrul sau se ia o decizie în urma parcurgerii unui test statistic.

## 3. Repartiţii bidimensionale

Considerăm două variabile aleatoare $X$ şi $Y$. Pe baza lor se defineşte vectorul aleator: $V: (X,Y)$.

În funcţie de tipul celor două variabile aleatoare, distingem două tipuri de vectori aleatori sau două cazuri: discret şi continuu.

- cazul discret:

$$V:\begin{pmatrix}(x_i,y_j)\\ p_{ij}\end{pmatrix}_{i\in I,\,j\in J},\quad p_{ij}=P[(X=x_i)\cap(Y=y_j)],$$

$$\sum_i\sum_j p_{ij}=1$$

- cazul continuu:

$$V: f(x,y),\quad f(x,y)\geq 0$$

$$(\forall)(x,y)\in R^2,\quad \int_R\int_R f(x,y)dxdy=1$$

$$F(x,y)=P(X<x,Y<y),\quad F(x,y)=\int_{-\infty}^{x}\int_{-\infty}^{y} f(s,t)dsdt$$

$$f(x,y)=\frac{\partial^2 F}{\partial x\partial y}$$

BCU IAŞI - UNIVERSITY LIBRARY

## Repartiţii marginale

- cazul discret:

$$X:\begin{pmatrix} x_i \\ p_{i.} \end{pmatrix}_{i\in I},\ p_{i.}=\sum_j p_{ij},\quad \sum_i p_{i.}=1$$

$$Y:\begin{pmatrix} y_j \\ p_{.j} \end{pmatrix}_{j\in J},\ p_{.j}=\sum_i p_{ij},\quad \sum_j p_{.j}=1$$

$$M(X)=\sum_i x_i p_{i.}=\sum_i\sum_j x_i p_{ij},\quad M(Y)=\sum_j y_j p_{.j}=\sum_i\sum_j y_j p_{ij}$$

- cazul continuu:

$$X: f_1(x)=\int_R f(x,y)dy,\quad F_1(x)=\int_{-\infty}^{x}\int_{-\infty}^{\infty} f(s,t)dsdt=\int_{-\infty}^{x} f_1(s)ds$$

$$Y: f_2(y)=\int_R f(x,y)dx,\quad F_2(y)=\int_{-\infty}^{y}\int_{-\infty}^{\infty} f(s,t)dsdt=\int_{-\infty}^{y} f_2(s)ds$$

$$M(X)=\int_R x f_1(x)dx=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x f(x,y)dxdy$$

## Repartiţii condiţionate

- cazul discret:

$$X/Y=y_j:\begin{pmatrix} x_i \\ \dfrac{p_{ij}}{p_{.j}} \end{pmatrix}_{i\in I},\ P(X=x_i/Y=y_j)=\frac{p_{ij}}{p_{.j}}$$

$$Y/X=x_i:\begin{pmatrix} y_j \\ \dfrac{p_{ij}}{p_{i.}} \end{pmatrix}_{j\in J},\ P(Y=y_j/X=x_i)=\frac{p_{ij}}{p_{i.}}$$

- cazul continuu:

$$X/Y: f(x/y)=\frac{f(x,y)}{f_2(y)},\quad f_2(y)\neq 0$$

$$Y/X: f(y/x)=\frac{f(x,y)}{f_1(x)},\quad f_1(x)\neq 0$$

BCU IAŞI CENTRAL UNIVERSITY LIBRARY

### Matricea de covarianţă

Pentru două variabile aleatoare $X$ şi $Y$, matricea de covarianţă se defineşte prin relaţia de mai jos:

$$\begin{pmatrix} V(X) & cov(X,Y) \\ cov(X,Y) & V(Y) \end{pmatrix}$$

### Medii condiţionate

- cazul discret:

$$M(X/Y=y_j)=\sum_i x_i \frac{p_{ij}}{p_{.j}}=\frac{1}{p_{.j}}\sum_i x_i p_{ij}=\alpha_j,\ M(X/Y)=\alpha(Y)$$

$$M(Y/X=x_i)=\sum_j y_j \frac{p_{ij}}{p_{i.}}=\frac{1}{p_{i.}}\sum_j y_j p_{ij}=\beta_i,\ M(Y/X)=\beta(X)$$

- cazul continuu:

$$M(X/Y)=\int_R xf(x/y)dx=\int_R x\frac{f(x,y)}{f_2(y)}dx=\alpha(Y)$$

$$M(Y/X)=\int_R yf(y/x)dy=\int_R y\frac{f(x,y)}{f_1(x)}dy=\beta(X)$$

## 4. Principalele repartiţii de probabilitate

### A. Repartiţii discrete

Dintre repartiţiile discrete, prezentăm: repartiţia Bernoulli, binomială şi Poisson.

### 1. Repartiţia Bernoulli

Repartiţia Bernoulli se referă la variabilele aleatoare cele mai simple, care pot lua doar două valori posibile, în funcţie de realizarea sau nu a unei proprietăţi.

Repartiţia este dată printr-un tablou de forma:

$$X:\begin{pmatrix} 0 & 1 \\ q & p \end{pmatrix}, \text{ unde}$$

$p=P(X=1)$, $q=P(X=0)$ şi $p+q=1$.

Dacă o variabilă aleatoare urmează o lege de repartiție Bernoulli, se simbolizează prin $X \sim B(p)$. Parametrul $p$ reprezintă media sau speranța matematică a variabilei de tip Bernoulli şi reprezintă probabilitatea de a se realiza proprietatea cerută sau de a apare evenimentul dorit, favorabil.

Parametrii acestei repartiții sunt:

$$M(X) = p$$
$$V(X) = pq$$

**Exemplu.** Pentru valoarea $p = 0,2$, repartiția Bernoulii se prezintă în figura de mai jos.

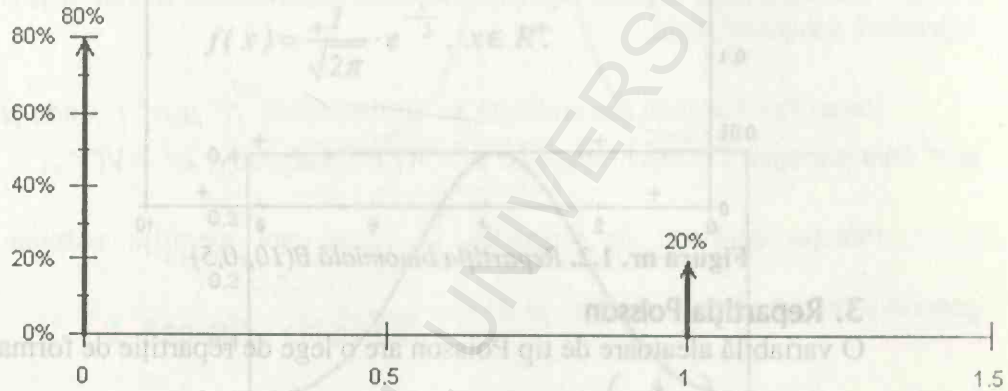



**Figura nr. 1.1.** *Repartiția Bernoulli B(0,2)*

## 2. Repartiţia binomială

Repartiția binomială se obține prin generalizarea repartiției Bernoulli. Prin însumarea unui număr de $n$ variabile aleatoare Bernoulli identic repartizate, se obține o variabilă binomială.

Repartiția binomială, simbolizată $B(n, p)$, este dată printr-un tablou de forma:

$$X : \begin{pmatrix} k \\ C_n^k p^k q^{n-k} \end{pmatrix}_{k=\overline{0,n}}$$, unde $p + q = 1$, iar $k$ reprezintă numărul de realizări a evenimentului favorabil în condițiile repetării de $n$ ori a experienței Bernoulli.

Parametrii repartiției binomiale se determină uşor, dacă se consideră că o variabilă binomială este suma a $n$ variabile Bernoulli, $X_i \sim B(p)$.

Considerăm variabila binomială $X = \sum_{i=1}^{n} X_i$, pentru care au loc relațiile:

$M(X) = np$
$V(X) = npq$

**Exemplu.** Pentru o experienţă Bernoulli repetată de 10 ori şi o probabilitate de 0,5 se obţine următoarea repartiţie binomială:

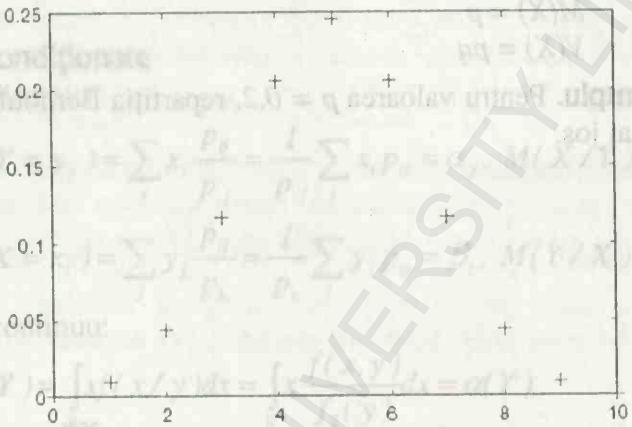



**Figura nr. 1.2.** *Repartiţia binomială B(10; 0,5)*

## 3. Repartiţia Poisson

O variabilă aleatoare de tip Poisson are o lege de repartiţie de forma:

$$X : \left( \frac{e^{-\lambda}\lambda^k}{k!} \right)_{k=\overline{0,n}}$$, unde $k$ reprezintă numărul de apariţii a unui anumit eveniment într-o anumită perioadă de timp.

Specific repartiţiei Poisson, simbolizată $P(\lambda)$, este că media şi varianţa sunt egale şi au valoarea: $M(X) = V(X) = \lambda$.

**Exemplu.** Pentru $\lambda = 5$, se obţine o repartiţie Poisson prezentată în figura de mai jos.

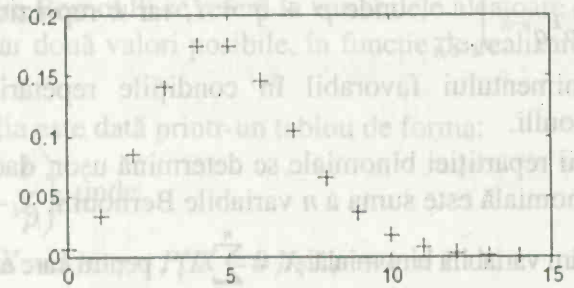



**Figura nr. 1.3.** *Repartiţia Poisson P(5)*

BCU IASI - CENTRAL UNIVERSITY LIBRARY

## B. Repartiţii continue

Dintre repartiţiile continue, prezentăm: normală standard şi generalizată, Student, chi-pătrat şi Fisher.

### 1. Repartiţia normală

#### Repartiţia normală standard

O variabilă aleatoare repartizată după o lege normală standard, simbolizată *N(0,1)*, are o funcţie densitate de repartiţie dată de relaţia:

$$f(x)=\frac{1}{\sqrt{2\pi}}\cdot e^{-\frac{x^2}{2}},\ x\in R.$$

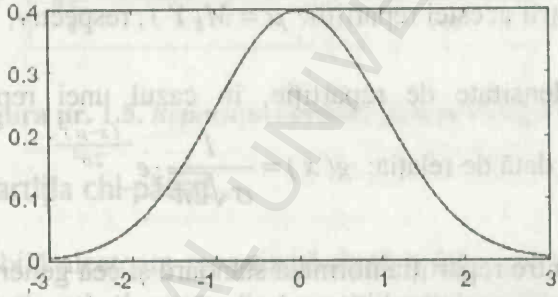



**Figura nr. 1.4.** *Repartiţia normală standard N(0,1)*

Parametrii repartiţiei normale standard sunt:
$M(X) = 0$
$V(X) = 1$

Pentru interese practice, de calcul a unor probabilităţi, se utilizează funcţia lui Laplace, definită pe baza repartiţiei normale standard. Funcţia lui Laplace este definită de relaţia:

$\Phi(x)=\int_0^x \frac{1}{\sqrt{2\pi}}\cdot e^{-\frac{t^2}{2}}dt$, iar valorile acestei funcţii se găsesc tabelate.

Pe baza funcţiei lui Laplace se poate determina, de exemplu, probabilitatea ca variabila aleatoare normală standard să ia valori într-un interval simetric de tipul *(-a; a)*:

$$P(-a<X<a)=F(a)-F(-a)=\Phi(a)-\Phi(-a)=2\Phi(a)=\int_{-a}^{a} f(t)dt.$$

**Exemplu.** Pentru $x = 1$, se poate calcula probabilitatea $P(-1 < X < 1) = 2\,\Phi(1) = 0{,}682$.

### Repartiţia normală generalizată

Dacă $X$ este o variabilă aleatoare normală standard, atunci variabila $Y = \mu + \sigma X$ urmează o lege de repartiţie normală generalizată, în care $\mu$ şi $\sigma$ reprezintă constante reale.

Repartiţia normală generalizată se simbolizează $N(\mu,\sigma^2)$, unde $\mu$ şi $\sigma^2$ sunt parametrii acestei repartiţii: $\mu = M(Y)$, respectiv, $\sigma^2 = V(Y)$.

Funcţia densitate de repartiţie, în cazul unei repartiţii normale generalizate, este dată de relaţia: $g(x)=\frac{1}{\sigma\sqrt{2\pi}}\cdot e^{-\frac{(x-\mu)^2}{2\sigma^2}}$.

Relaţia dintre repartiţia normală standard şi cea generalizată se poate observa uşor atât prin relaţia dintre cele două tipuri de variabile, cât şi prin relaţia existentă între funcţiile de repartiţie specifice. Astfel, unei variabile normale standard $X$, cu o funcţie densitate $f(x)$ şi o funcţie de repartiţie $F(x)$ îi corespunde o variabilă normală generalizată $Y = \mu + \sigma X$, cu o funcţie densitate $g(x)$ şi o funcţie de repartiţie $G(x)$, unde:

$$G(x)=F\left(\frac{x-\mu}{\sigma}\right).$$

Invers, unei variabile normale generalizate $Y$ îi corespunde o variabilă normală standard $X=\frac{Y-\mu}{\sigma}$. Ultima relaţie poartă numele de relaţie de *standardizare a unei variabile aleatoare*.

În baza operaţiei de standardizare şi a relaţiei dintre cele două funcţii de repartiţie, se poate utiliza funcţia lui Laplace pentru a calcula anumite probabilităţi pentru o repartiţie normală generalizată. Astfel, de exemplu, se poate calcula probabilitatea:

$$P(a<Y<b)=G(b)-G(a)=F(\frac{b-\mu}{\sigma})-F(\frac{a-\mu}{\sigma})=\Phi(\frac{b-\mu}{\sigma})-\Phi(\frac{a-\mu}{\sigma})$$

**Exemplu.** O variabilă aleatoare normală generalizată de parametri *10* şi *2* se prezintă în figura de mai jos.

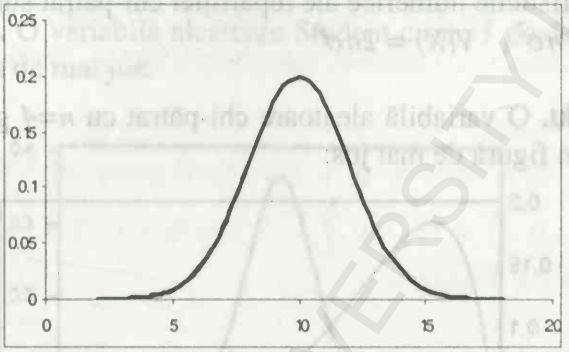



**Figura nr. 1.5.** *Repartiţia normală generalizată N(10,2)*

## 2. Repartiţia chi-pătrat

O variabilă aleatoare repartizată după o lege chi-pătrat, simbolizată $\chi^2(n,\sigma)$, are o funcţie densitate de repartiţie dată prin relaţia:

$$f(x)=\begin{cases}\dfrac{1}{2^{\frac{n}{2}}\sigma^n\Gamma\left(\dfrac{n}{2}\right)}\cdot x^{\frac{n}{2}-1}e^{-\frac{x}{2\sigma^2}}, & x>0\\ 0, & x\le 0\end{cases}, \text{ unde}$$

$\Gamma(a)=\int_0^{\infty}x^{a-1}e^{-x}dx, \quad a>0$ este integrala Euler.

Dacă considerăm *n* variabile aleatoare identic repartizate după o lege normală standard, $X_i \sim N(0,1)$, atunci variabila $X=\sum_{i=1}^{n}X_i^2 \sim \chi^2(n)$. Cu alte cuvinte, suma pătratelor a *n* variabile aleatoare normal repartizate reprezintă o variabilă aleatoare repartizată chi-pătrat.

Parametrii repartiției chi-pătrat sunt determinați de modul de construire a variabilei. Astfel, $n$ reprezintă *numărul de grade de libertate*, adică numărul de variabile aleatoare normale la pătrat însumate, iar $\sigma$ reprezintă abaterea standard a variabilei aleatoare normale (care este egală cu unu în cazul variabilei standard).

Caracteristicile numerice ale repartiției chi-pătrat sunt:
$M(X) = n\sigma^2$, $V(X) = 2n\sigma^4$

**Exemplu.** O variabilă aleatoare chi-pătrat cu $n=4$ grade de libertate se prezintă ca în figura de mai jos:

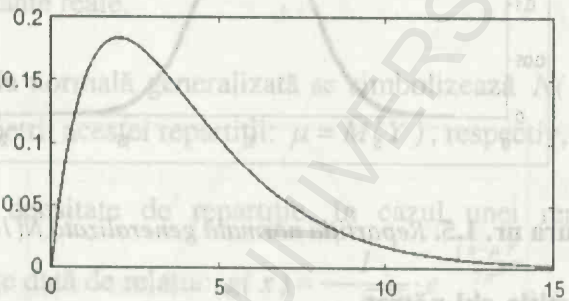



**Figura nr. 1.6.** *Repartiția chi-pătrat* $\chi^2(4)$

## 3. Repartiția Student

O variabilă aleatoare repartizată după o lege Student, simbolizată $t(n)$, are o funcție densitate de repartiție dată de relația:

$$f(x) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi}\cdot\Gamma\left(\frac{n}{2}\right)}\cdot\left(1+\frac{x^2}{n}\right)^{-\frac{n+1}{2}}, \quad x\in R, \text{ unde}$$

$\Gamma(a) = \int_0^{\infty} x^{a-1}e^{-x}dx$, $a>0$ este integrala Euler.

Dacă se consideră două variabile aleatoare $X \sim N(0,1)$ și $Y \sim \chi^2(n)$, atunci variabila aleatoare student se obține prin relația: $t = \frac{X}{\sqrt{\frac{Y}{n}}} \sim t(n)$, unde $n$ reprezintă numărul de grade de libertate, parametrul repartiției Student.

**Exemplu.** O variabilă aleatoare Student cu $n=3$ grade de libertate se prezintă în figura de mai jos:

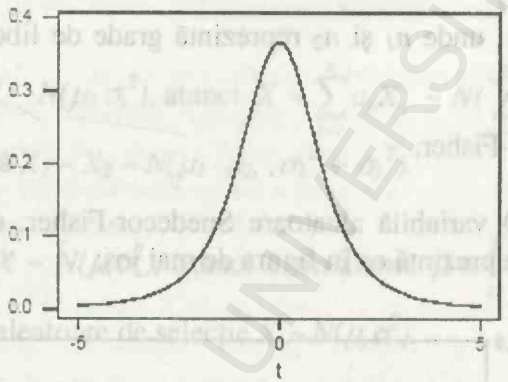



**Figura nr. 1.7.** *Repartiția Student t(3)*

## 4. Repartiția Snedecor-Fisher

O variabilă aleatoare repartizată după o lege Snedecor-Fisher, simbolizată $\mathcal{S}(n_1, n_2)$, are o funcție densitate de repartiție dată prin relația:

$$f(x) = \begin{cases} \dfrac{\left(\dfrac{n_1}{n_2}\right)^{\frac{n_1}{2}}}{B\left(\dfrac{n_1}{2}, \dfrac{n_2}{2}\right)} \cdot x^{\frac{n_1}{2}-1} \left(1+\dfrac{n_1}{n_2}x\right)^{-\frac{n_1+n_2}{2}}, & x>0, \text{ unde} \\ 0, & x \le 0 \end{cases}$$

$B(a,b) = \int_0^1 x^{a-1}(1-x)^{b-1}dx, \quad a>0, b>0$ este integrala Euler.

Caracteristicile numerice ale repartiției Sedecor-Fisher sunt:

$$M(X) = \frac{n_2}{n_2 - 2}, \quad n_2 > 2$$

$$V(X) = \frac{2n_2^2(n_1 + n_2 - 2)}{n_1(n_2 - 2)^2(n_2 - 4)}, \quad n_2 > 4$$

Dacă se consideră două variabile aleatoare: $X \sim \chi^2(n_1, \sigma)$ și $Y \sim \chi^2(n_2, \sigma)$, atunci o variabilă repartizată Fisher se obține prin relația: $F = \frac{\frac{X}{n_1}}{\frac{Y}{n_2}} \sim \Im(n_1, n_2)$, unde $n_1$ și $n_2$ reprezintă grade de libertate, parametrii repartiției Snedecor-Fisher.

**Exemplu.** O variabilă aleatoare Snedecor-Fisher, cu $n_1=5$ și $n_2=3$ grade de libertate se prezintă ca în figura de mai jos:

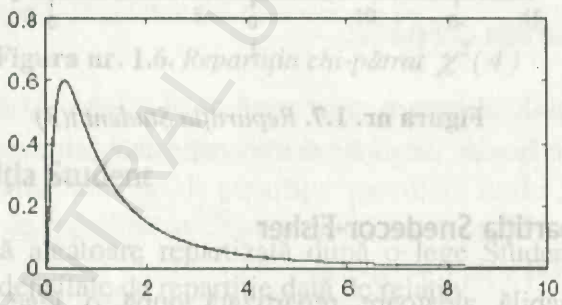



**Figura nr. 1.8.** *Repartiția Snedecor-Fisher F(5, 3)*

## 5. Câteva rezultate în cazul eşantioanelor extrase din colectivităţi normale

Considerăm următoarele rezultate utilizate mai des în problemele de inferență statistică, pentru eşantioane extrase din colectivități normal distribuite.

**1.** Dacă $Y \sim N(\mu, \sigma^2)$, atunci $X = \frac{Y - \mu}{\sigma} \sim N(0,1)$.

BCU IASI CENTRAL UNIVERSITY LIBRARY

**2.** Dacă $X_i \sim N(0,\sigma^2)$, atunci $X = \sum_{i=1}^{n} X_i^2 \sim \chi^2(n,\sigma)$.

**3.** Dacă $X_i \sim \chi^2(n_i,\sigma)$, atunci $X = \sum_{i=1}^{k} X_i \sim \chi^2(\sum_{i=1}^{k} n_i,\sigma)$.

**4.** Dacă $X \sim N(0, \sigma^2)$ şi $Y \sim \chi^2(n,\sigma)$, atunci $t = \frac{X}{\sqrt{\frac{Y}{n}}} \sim t(n)$.

**5.** Dacă $X_i \sim N(\mu_i,\sigma_i^2)$, atunci $X = \sum_{i=1}^{n} a_i X_i \sim N(\sum_{i=1}^{n} a_i\mu_i, \sum_{i=1}^{n} a_i^2\sigma_i^2)$.

Rezultă că $X_1 - X_2 \sim N(\mu_1 - \mu_2, , \sigma_1^2 + \sigma_2^2)$.

**6.** Dacă $X \sim N(\mu,\sigma^2)$, atunci estimatorul $\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} X_i \sim N(\mu, \frac{\sigma^2}{n})$, unde variabilele aleatoare de selecţie $X_i \sim N(\mu,\sigma^2)$.

**7.** Dacă $X \sim N(\mu,\sigma^2)$, atunci estimatorul standardizat este $\frac{\hat{\mu} - \mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0,1)$.

**8.** Dacă $X_1 \sim N(\mu_1,\sigma_1^2)$, $X_2 \sim N(\mu_2,\sigma_2^2)$, atunci $\hat{\mu}_1 - \hat{\mu}_2 \sim N(\mu_1 - \mu_2, \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2})$.

**9.** Dacă $X \sim N(\mu,\sigma^2)$, atunci $Y = \sum_{i=1}^{n} \left(\frac{X_i - \mu}{\sigma}\right)^2 \sim \chi^2(n,1)$.

**10.** Dacă $X \sim N(\mu,\sigma^2)$, atunci $Y = \frac{(n-1)}{\sigma^2}\hat{\sigma}'^2 \sim \chi^2(n-1,1)$,

unde $\hat{\sigma}'^2$ reprezintă dispersia de selecţie modificată, estimatorul nedeplasat pentru parametrul $\sigma^2$.

BCU IAŞI - CENTRAL UNIVERSITY LIBRARY

**11.** Dacă $X \sim N(\mu, \sigma^2)$, atunci $t = \dfrac{\hat{\mu} - \mu}{\sqrt{\dfrac{\hat{\sigma}'^2}{n}}} \sim t(n-1)$.

**12.** Dacă $X \sim N(\mu, \sigma^2)$, atunci $t = \dfrac{\hat{\mu} - \mu}{\sqrt{\dfrac{\hat{\sigma}^2}{n-1}}} \sim t(n-1)$,

unde $\hat{\sigma}^2$ este dispersia de selecție, estimatorul deplasat pentru parametrul $\sigma^2$.

**13.** Dacă $X_1 \sim N(\mu_1, \sigma_1^2)$, $X_2 \sim N(\mu_2, \sigma_2^2)$, atunci

$$t = \sqrt{\frac{n_1 n_2 (n_1 + n_2 - 2)}{n_1 + n_2}} \cdot \frac{\hat{\mu}_1 - \hat{\mu}_2 - (\mu_1 - \mu_2)}{\sqrt{(n_1 - 1)\hat{\sigma}'^2_1 + (n_2 - 1)\hat{\sigma}'^2_2}} \sim t(n_1 + n_2 - 2).$$

**14.** Dacă $X_1 \sim N(\mu_1, \sigma_1^2)$, $X_2 \sim N(\mu_2, \sigma_2^2)$, atunci

$$F = \frac{\sigma_2^2}{\sigma_1^2} \frac{\hat{\sigma}'^2_1}{\hat{\sigma}'^2_2} \sim \mathcal{F}(n_1, n_2).$$

BCU IASI/CENTRAL UNIVERSITY LIBRARY

# Bibliografie


1. Agresti, A. – *Categorical data analysis,* John Wiley and Sons, 1990
2. Berdot, J.P. – *Econometrie* – CNED, Poitiers-Futurscope, 2001
3. Berrebi, L. – *Estimation par regression simple*, INSEE – Methodes, Nr. 29-30-31, dec. 1992, pp. 239-264
4. Biji, M., Biji, M.E., Lilea, E., Anghelache, C. – *Tratat de statistică,* Editura Economică, Bucureşti, 2005
5. Blaug, M. – *Teoria economică în retrospectivă*, Editura Didactică şi Pedagogică, Bucureşti, 1992
6. Bourbonnais, R. – *Econometrie,* Dunod, Paris, 2000
7. Caperaa, P., Van Cutsem, B. – *Méthodes et modèles en statistique non parametrique*, Dunod, Paris, 1988
8. Daudin, J., Robin, S. – *Statistique inférentielle*, Presses Universitaires de Rennes, 1999
9. Deutsch, R. – *Estimation Theory*, Prentice-Hall, Inc., New Jersey, 1965
10. Dumas de Rauly, D. – *L'estimation statistique*, Gauthier-Villars, Paris, 1966
11. Eatwell, J., Newman, P. – *Times series and statistics*, W.W. Norton, New York, 1996
12. Fisher, R.A. – *Les méthodes statistiques*, Presses Universitaires de France, Paris, 1947
13. Gourieroux, C., Monfort, A. –*Statistique et modèles économetriques*, Economica, Paris, 1989, 2vol.
14. Gujarati, D.N. – *Basic econometrics*, McGraw-Hill, New York, 1995
15. Isaic-Maniu, Al., Mitruţ, C., Voineagu, V. – *Statistica*, Editura Universitară, Bucureşti, 2004
16. Iacob, A. I., Tanasoiu, O. – *Modele econometrice* : vol. I , Editura ASE, Bucureşti, 2005
17. Jaba, E., Niculicioiu, P., Bilaus, M. – *Cercetarea selectivă*, Universitatea „Al.I. Cuza", Iaşi, 1977
18. Jaba, Elisabeta, *Statistica economiei naţionale*, Universitatea „Al.I. Cuza", Iaşi, 1982
19. Jaba, Elisabeta – *Statistica*, ediţia a treia, Editura Economică, Bucureşti, 2002

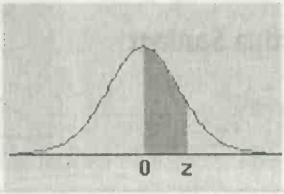

BCU IASI / CENTRAL UNIVERSITY LIBRARY

20. Jaba, E., Grama, A. – *Analiza statistică cu SPSS sub Windows,* Polirom, Iaşi, 2004
21. Jaba, E., Pintilescu, C. – *Statistică. Teste grilă şi probleme,* Sedcom Libris, Iaşi, 2005
22. Jaba, E., Pintilescu, C., Jemna, D.V. – *Statistică inferenţială. Teste grilă şi probleme,* Sedcom Libris, Iaşi, 2002
23. Jaba, Elisabeta – *Note curs econometrie,* 2006
24. Jemna, D.V. – *Eficienţa sondajului statistic,* Sedcom Libris, Iaşi, 2005
25. Jemna, D.V. – *Note curs econometrie,* 2005, 2006
26. Korka, M., Begu, L., Roşca E., Tuşa E. - *Aplicaţii statistice,* Editura Bucovina Viitoare, Suceava, 2002
27. Johnston, J., Dinardo, J. – *Méthodes économetriques*, Economica, Paris, 1997
28. Kendall, M.G., Stuart, A. – *The Advanced Theory of Statistics*, Charles Griffin, London, 1963, 3 vol.
29. Malinvaud, E. – *Méthodes statistiques de l'économetrie,* Dunod, Paris, 1978
30. McEachern, W.A. – *Economics,* South-Western College Publishing, 2000
31. Mihoc, G., Craiu, V. – *Tratat de statistică matematică*, volumul I, Editura Academiei R.S.R., Bucureşti, 1976
32. Monfort, A. – *Cours de statistique mathématique*, Economica, Paris, 1997
33. Nenciu, E. – *Teoria probabilităţilor şi statistică matematică*, Editura Universităţii „Al. I. Cuza" Iaşi, 1984
34. Onicescu, Octav – *Principes de logique et de philosophie mathématique*, Bucureşti, Editura Academiei RSR, 1971
35. Pârţachi, I., Brăilă, A., Şişcanu, N. – *Economerie aplicată,* ASEM, Chişinău, 1999
36. Pecican, E.S. – *Econometria pentru economişti,* Ed. Economică, Bucureşti, 2003
37. Pintilescu, C. – *Analiza datelor,* Junimea, Iaşi, 2002
38. Tassi, P. – *Méthodes statistiques*, Economica, Paris, 1989
39. Tiganescu, E., Roman, M. – *Macroeconomie: o abordare cantitativă*, Editura Economica, 2005
40. Wilthien, P.H. – *Décision statistique et économetrie*, Armand Colin, Paris, 1996
41. Yule, U.G., Kendall, M.C. – *Introducere în teoria statisticii*, Editura Ştiinţifică, Bucureşti, 1969
42. Zaiţ, D., Nica, P. – *Introducere în modelarea econometrică,* Editura Universităţii „Al.I. Cuza", Iaşi, 1995

BCU IASI / CENTRAL UNIVERSITY LIBRARY

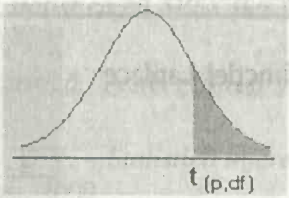
# Tabele probabiliste

## Valorile funcţiei Laplace



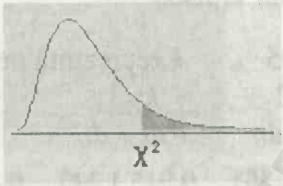

| | 0.00 | 0.01 | 0.02 | 0.03 | 0.04 | 0.05 | 0.06 | 0.07 | 0.08 | 0.09 |
|---|---|---|---|---|---|---|---|---|---|---|
| 0.0 | 0.0000 | 0.0040 | 0.0080 | 0.0120 | 0.0160 | 0.0199 | 0.0239 | 0.0279 | 0.0319 | 0.0359 |
| 0.1 | 0.0398 | 0.0438 | 0.0478 | 0.0517 | 0.0557 | 0.0596 | 0.0636 | 0.0675 | 0.0714 | 0.0753 |
| 0.2 | 0.0793 | 0.0832 | 0.0871 | 0.0910 | 0.0948 | 0.0987 | 0.1026 | 0.1064 | 0.1103 | 0.1141 |
| 0.3 | 0.1179 | 0.1217 | 0.1255 | 0.1293 | 0.1331 | 0.1368 | 0.1406 | 0.1443 | 0.1480 | 0.1517 |
| 0.4 | 0.1554 | 0.1591 | 0.1628 | 0.1664 | 0.1700 | 0.1736 | 0.1772 | 0.1808 | 0.1844 | 0.1879 |
| 0.5 | 0.1915 | 0.1950 | 0.1985 | 0.2019 | 0.2054 | 0.2088 | 0.2123 | 0.2157 | 0.2190 | 0.2224 |
| 0.6 | 0.2257 | 0.2291 | 0.2324 | 0.2357 | 0.2389 | 0.2422 | 0.2454 | 0.2486 | 0.2517 | 0.2549 |
| 0.7 | 0.2580 | 0.2611 | 0.2642 | 0.2673 | 0.2704 | 0.2734 | 0.2764 | 0.2794 | 0.2823 | 0.2852 |
| 0.8 | 0.2881 | 0.2910 | 0.2939 | 0.2967 | 0.2995 | 0.3023 | 0.3051 | 0.3078 | 0.3106 | 0.3133 |
| 0.9 | 0.3159 | 0.3186 | 0.3212 | 0.3238 | 0.3264 | 0.3289 | 0.3315 | 0.3340 | 0.3365 | 0.3389 |
| 1.0 | 0.3413 | 0.3438 | 0.3461 | 0.3485 | 0.3508 | 0.3531 | 0.3554 | 0.3577 | 0.3599 | 0.3621 |
| 1.1 | 0.3643 | 0.3665 | 0.3686 | 0.3708 | 0.3729 | 0.3749 | 0.3770 | 0.3790 | 0.3810 | 0.3830 |
| 1.2 | 0.3849 | 0.3869 | 0.3888 | 0.3907 | 0.3925 | 0.3944 | 0.3962 | 0.3980 | 0.3997 | 0.4015 |
| 1.3 | 0.4032 | 0.4049 | 0.4066 | 0.4082 | 0.4099 | 0.4115 | 0.4131 | 0.4147 | 0.4162 | 0.4177 |
| 1.4 | 0.4192 | 0.4207 | 0.4222 | 0.4236 | 0.4251 | 0.4265 | 0.4279 | 0.4292 | 0.4306 | 0.4319 |
| 1.5 | 0.4332 | 0.4345 | 0.4357 | 0.4370 | 0.4382 | 0.4394 | 0.4406 | 0.4418 | 0.4429 | 0.4441 |
| 1.6 | 0.4452 | 0.4463 | 0.4474 | 0.4484 | 0.4495 | 0.4505 | 0.4515 | 0.4525 | 0.4535 | 0.4545 |
| 1.7 | 0.4554 | 0.4564 | 0.4573 | 0.4582 | 0.4591 | 0.4599 | 0.4608 | 0.4616 | 0.4625 | 0.4633 |
| 1.8 | 0.4641 | 0.4649 | 0.4656 | 0.4664 | 0.4671 | 0.4678 | 0.4686 | 0.4693 | 0.4699 | 0.4706 |
| 1.9 | 0.4713 | 0.4719 | 0.4726 | 0.4732 | 0.4738 | 0.4744 | 0.4750 | 0.4756 | 0.4761 | 0.4767 |
| 2.0 | 0.4772 | 0.4778 | 0.4783 | 0.4788 | 0.4793 | 0.4798 | 0.4803 | 0.4808 | 0.4812 | 0.4817 |
| 2.1 | 0.4821 | 0.4826 | 0.4830 | 0.4834 | 0.4838 | 0.4842 | 0.4846 | 0.4850 | 0.4854 | 0.4857 |
| 2.2 | 0.4861 | 0.4864 | 0.4868 | 0.4871 | 0.4875 | 0.4878 | 0.4881 | 0.4884 | 0.4887 | 0.4890 |
| 2.3 | 0.4893 | 0.4896 | 0.4898 | 0.4901 | 0.4904 | 0.4906 | 0.4909 | 0.4911 | 0.4913 | 0.4916 |
| 2.4 | 0.4918 | 0.4920 | 0.4922 | 0.4925 | 0.4927 | 0.4929 | 0.4931 | 0.4932 | 0.4934 | 0.4936 |
| 2.5 | 0.4938 | 0.4940 | 0.4941 | 0.4943 | 0.4945 | 0.4946 | 0.4948 | 0.4949 | 0.4951 | 0.4952 |
| 2.6 | 0.4953 | 0.4955 | 0.4956 | 0.4957 | 0.4959 | 0.4960 | 0.4961 | 0.4962 | 0.4963 | 0.4964 |
| 2.7 | 0.4965 | 0.4966 | 0.4967 | 0.4968 | 0.4969 | 0.4970 | 0.4971 | 0.4972 | 0.4973 | 0.4974 |
| 2.8 | 0.4974 | 0.4975 | 0.4976 | 0.4977 | 0.4977 | 0.4978 | 0.4979 | 0.4979 | 0.4980 | 0.4981 |
| 2.9 | 0.4981 | 0.4982 | 0.4982 | 0.4983 | 0.4984 | 0.4984 | 0.4985 | 0.4985 | 0.4986 | 0.4986 |
| 3.0 | 0.4987 | 0.4987 | 0.4987 | 0.4988 | 0.4988 | 0.4989 | 0.4989 | 0.4989 | 0.4990 | 0.4990 |

## Repartiţia Student




| df\p | 0.10 | 0.05 | 0.025 | 0.01 | 0.005 |
|---|---|---|---|---|---|
| 1 | 3,078 | 6,314 | 12,706 | 31,821 | 63,657 |
| 2 | 1,886 | 2,920 | 4,303 | 6,965 | 9,925 |
| 3 | 1,638 | 2,353 | 3,182 | 4,541 | 5,841 |
| 4 | 1,533 | 2,132 | 2,776 | 3,747 | 4,604 |
| 5 | 1,476 | 2,015 | 2,571 | 3,365 | 4,032 |
| 6 | 1,440 | 1,943 | 2,447 | 3,143 | 3,707 |
| 7 | 1,415 | 1,895 | 2,365 | 2,998 | 3,499 |
| 8 | 1,397 | 1,860 | 2,306 | 2,896 | 3,355 |
| 9 | 1,383 | 1,833 | 2,262 | 2,821 | 3,250 |
| 10 | 1,372 | 1,812 | 2,228 | 2,764 | 3,169 |
| 11 | 1,363 | 1,796 | 2,201 | 2,718 | 3,106 |
| 12 | 1,356 | 1,782 | 2,179 | 2,681 | 3,055 |
| 13 | 1,350 | 1,771 | 2,160 | 2,650 | 3,012 |
| 14 | 1,345 | 1,761 | 2,145 | 2,624 | 2,977 |
| 15 | 1,341 | 1,753 | 2,131 | 2,602 | 2,947 |
| 16 | 1,337 | 1,746 | 2,120 | 2,583 | 2,921 |
| 17 | 1,333 | 1,740 | 2,110 | 2,567 | 2,898 |
| 18 | 1,330 | 1,734 | 2,101 | 2,552 | 2,878 |
| 19 | 1,328 | 1,729 | 2,093 | 2,539 | 2,861 |
| 20 | 1,325 | 1,725 | 2,086 | 2,528 | 2,845 |
| 21 | 1,323 | 1,721 | 2,080 | 2,518 | 2,831 |
| 22 | 1,321 | 1,717 | 2,074 | 2,508 | 2,819 |
| 23 | 1,319 | 1,714 | 2,069 | 2,500 | 2,807 |
| 24 | 1,318 | 1,711 | 2,064 | 2,492 | 2,797 |
| 25 | 1,316 | 1,708 | 2,060 | 2,485 | 2,787 |
| 26 | 1,315 | 1,706 | 2,056 | 2,479 | 2,779 |
| 27 | 1,314 | 1,703 | 2,052 | 2,473 | 2,771 |
| 28 | 1,313 | 1,701 | 2,048 | 2,467 | 2,763 |
| 29 | 1,311 | 1,699 | 2,045 | 2,462 | 2,756 |
| 30 | 1,310 | 1,697 | 2,042 | 2,457 | 2,750 |
| inf | 1,282 | 1,645 | 1,960 | 2,326 | 2,576 |

## Repartiţia Chi-pătrat




| df\area | .100 | .050 | .025 | .010 | .005 |
|---|---|---|---|---|---|
| 1 | 2,706 | 3,841 | 5,024 | 6,635 | 7,879 |
| 2 | 4,605 | 5,991 | 7,378 | 9,210 | 10,597 |
| 3 | 6,251 | 7,815 | 9,348 | 11,345 | 12,838 |
| 4 | 7,779 | 9,488 | 11,143 | 13,277 | 14,860 |
| 5 | 9,236 | 11,071 | 12,833 | 15,086 | 16,750 |
| 6 | 10,645 | 12,592 | 14,449 | 16,812 | 18,548 |
| 7 | 12,017 | 14,067 | 16,013 | 18,475 | 20,278 |
| 8 | 13,362 | 15,507 | 17,535 | 20,090 | 21,955 |
| 9 | 14,684 | 16,919 | 19,023 | 21,666 | 23,589 |
| 10 | 15,987 | 18,307 | 20,483 | 23,209 | 25,188 |
| 11 | 17,275 | 19,675 | 21,920 | 24,725 | 26,757 |
| 12 | 18,549 | 21,026 | 23,337 | 26,217 | 28,300 |
| 13 | 19,812 | 22,362 | 24,736 | 27,688 | 29,819 |
| 14 | 21,064 | 23,685 | 26,119 | 29,141 | 31,319 |
| 15 | 22,307 | 24,996 | 27,488 | 30,578 | 32,801 |
| 16 | 23,542 | 26,296 | 28,845 | 32,000 | 34,267 |
| 17 | 24,769 | 27,587 | 30,191 | 33,409 | 35,718 |
| 18 | 25,989 | 28,869 | 31,526 | 34,805 | 37,156 |
| 19 | 27,204 | 30,144 | 32,852 | 36,191 | 38,582 |
| 20 | 28,412 | 31,410 | 34,170 | 37,566 | 39,997 |
| 21 | 29,615 | 32,671 | 35,479 | 38,932 | 41,401 |
| 22 | 30,813 | 33,924 | 36,781 | 40,289 | 42,796 |
| 23 | 32,007 | 35,172 | 38,076 | 41,638 | 44,181 |
| 24 | 33,196 | 36,415 | 39,364 | 42,980 | 45,559 |
| 25 | 34,382 | 37,652 | 40,646 | 44,314 | 46,928 |
| 26 | 35,563 | 38,885 | 41,923 | 45,642 | 48,290 |
| 27 | 36,741 | 40,113 | 43,195 | 46,963 | 49,645 |
| 28 | 37,916 | 41,337 | 44,461 | 48,278 | 50,993 |
| 29 | 39,087 | 42,557 | 45,722 | 49,588 | 52,336 |
| 30 | 40,256 | 43,773 | 46,979 | 50,892 | 53,672 |

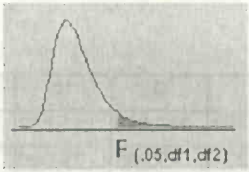
## Repartiţia Durbin-Watson

$\alpha = 0{,}05$ $k$ reprezintă numărul de parametri din model

| $n$ | $k=2$ | | $k=3$ | | $k=4$ | | $k=5$ | | $k=6$ | |
|---|---|---|---|---|---|---|---|---|---|---|
| | $d_L$ | $d_U$ | $d_L$ | $d_U$ | $d_L$ | $d_U$ | $d_L$ | $d_U$ | $d_L$ | $d_U$ |
| 15 | 1.077 | 1.361 | 0.945 | 1.543 | 0.814 | 1.750 | 0.685 | 1.977 | 0.562 | 2.220 |
| 16 | 1.106 | 1.371 | 0.982 | 1.539 | 0.857 | 1.728 | 0.734 | 1.935 | 0.615 | 2.157 |
| 17 | 1.133 | 1.381 | 1.015 | 1.536 | 0.897 | 1.710 | 0.779 | 1.900 | 0.664 | 2.104 |
| 18 | 1.158 | 1.392 | 1.046 | 1.535 | 0.933 | 1.696 | 0.820 | 1.872 | 0.710 | 2.060 |
| 19 | 1.180 | 1.401 | 1.075 | 1.535 | 0.967 | 1.685 | 0.859 | 1.848 | 0.752 | 2.022 |
| 20 | 1.201 | 1.411 | 1.100 | 1.537 | 0.998 | 1.676 | 0.894 | 1.828 | 0.792 | 1.991 |
| 21 | 1.221 | 1.420 | 1.125 | 1.538 | 1.026 | 1.669 | 0.927 | 1.812 | 0.828 | 1.964 |
| 22 | 1.240 | 1.429 | 1.147 | 1.541 | 1.053 | 1.664 | 0.958 | 1.797 | 0.863 | 1.940 |
| 23 | 1.257 | 1.437 | 1.168 | 1.543 | 1.078 | 1.660 | 0.986 | 1.786 | 0.895 | 1.919 |
| 24 | 1.273 | 1.446 | 1.188 | 1.546 | 1.101 | 1.657 | 1.013 | 1.775 | 0.925 | 1.902 |
| 25 | 1.288 | 1.454 | 1.206 | 1.550 | 1.123 | 1.654 | 1.038 | 1.767 | 0.953 | 1.886 |
| 26 | 1.302 | 1.461 | 1.224 | 1.553 | 1.143 | 1.652 | 1.062 | 1.759 | 0.979 | 1.873 |
| 27 | 1.316 | 1.468 | 1.240 | 1.556 | 1.162 | 1.651 | 1.083 | 1.753 | 1.004 | 1.861 |
| 28 | 1.328 | 1.476 | 1.255 | 1.560 | 1.181 | 1.650 | 1.104 | 1.747 | 1.028 | 1.850 |
| 29 | 1.341 | 1.483 | 1.270 | 1.563 | 1.198 | 1.650 | 1.124 | 1.743 | 1.050 | 1.841 |
| 30 | 1.352 | 1.489 | 1.284 | 1.567 | 1.214 | 1.650 | 1.143 | 1.739 | 1.070 | 1.833 |
| 31 | 1.363 | 1.496 | 1.297 | 1.570 | 1.229 | 1.650 | 1.160 | 1.735 | 1.090 | 1.825 |
| 32 | 1.373 | 1.502 | 1.309 | 1.573 | 1.244 | 1.650 | 1.177 | 1.732 | 1.109 | 1.819 |
| 33 | 1.383 | 1.508 | 1.321 | 1.577 | 1.258 | 1.651 | 1.193 | 1.730 | 1.127 | 1.813 |
| 34 | 1.393 | 1.514 | 1.332 | 1.580 | 1.271 | 1.652 | 1.208 | 1.728 | 1.144 | 1.807 |
| 35 | 1.402 | 1.519 | 1.343 | 1.584 | 1.283 | 1.653 | 1.222 | 1.726 | 1.160 | 1.803 |
| 36 | 1.411 | 1.524 | 1.354 | 1.587 | 1.295 | 1.654 | 1.236 | 1.725 | 1.175 | 1.799 |
| 37 | 1.419 | 1.530 | 1.364 | 1.590 | 1.307 | 1.655 | 1.249 | 1.723 | 1.190 | 1.795 |
| 38 | 1.427 | 1.535 | 1.373 | 1.594 | 1.317 | 1.656 | 1.261 | 1.723 | 1.204 | 1.792 |
| 39 | 1.435 | 1.540 | 1.382 | 1.597 | 1.328 | 1.658 | 1.273 | 1.722 | 1.218 | 1.789 |
| 40 | 1.442 | 1.544 | 1.391 | 1.600 | 1.338 | 1.659 | 1.285 | 1.721 | 1.231 | 1.786 |
| 45 | 1.475 | 1.566 | 1.430 | 1.615 | 1.383 | 1.666 | 1.336 | 1.720 | 1.287 | 1.776 |
| 50 | 1.503 | 1.585 | 1.462 | 1.628 | 1.421 | 1.674 | 1.378 | 1.721 | 1.334 | 1.771 |
| 55 | 1.527 | 1.601 | 1.490 | 1.640 | 1.452 | 1.681 | 1.414 | 1.724 | 1.374 | 1.768 |
| 60 | 1.549 | 1.616 | 1.514 | 1.652 | 1.480 | 1.689 | 1.444 | 1.727 | 1.408 | 1.767 |
| 65 | 1.567 | 1.629 | 1.536 | 1.662 | 1.503 | 1.696 | 1.471 | 1.731 | 1.438 | 1.767 |
| 70 | 1.583 | 1.641 | 1.554 | 1.671 | 1.524 | 1.703 | 1.494 | 1.735 | 1.464 | 1.768 |
| 75 | 1.598 | 1.652 | 1.571 | 1.680 | 1.543 | 1.709 | 1.515 | 1.739 | 1.486 | 1.770 |
| 80 | 1.611 | 1.662 | 1.586 | 1.688 | 1.560 | 1.715 | 1.534 | 1.743 | 1.507 | 1.772 |
| 85 | 1.624 | 1.671 | 1.600 | 1.696 | 1.575 | 1.721 | 1.551 | 1.747 | 1.525 | 1.774 |
| 90 | 1.635 | 1.679 | 1.612 | 1.703 | 1.589 | 1.726 | 1.566 | 1.751 | 1.542 | 1.776 |
| 95 | 1.645 | 1.687 | 1.623 | 1.709 | 1.602 | 1.732 | 1.579 | 1.755 | 1.557 | 1.778 |
| 100 | 1.654 | 1.694 | 1.634 | 1.715 | 1.613 | 1.736 | 1.592 | 1.758 | 1.571 | 1.780 |

BCU LIBRARY

## Repartiţia Fisher



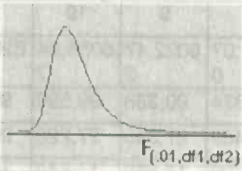

| df2/df1 | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|---|
| 1 | 161,448 | 199,500 | 215,707 | 224,583 | 230,162 | 233,986 | 236,768 |
| 2 | 18,513 | 19,000 | 19,164 | 19,247 | 19,296 | 19,330 | 19,353 |
| 3 | 10,128 | 9,552 | 9,277 | 9,117 | 9,014 | 8,941 | 8,887 |
| 4 | 7,709 | 6,944 | 6,591 | 6,388 | 6,256 | 6,163 | 6,094 |
| 5 | 6,608 | 5,786 | 5,410 | 5,192 | 5,050 | 4,950 | 4,876 |
| 6 | 5,987 | 5,143 | 4,757 | 4,534 | 4,387 | 4,284 | 4,207 |
| 7 | 5,591 | 4,737 | 4,347 | 4,120 | 3,972 | 3,866 | 3,787 |
| 8 | 5,318 | 4,459 | 4,066 | 3,838 | 3,688 | 3,581 | 3,501 |
| 9 | 5,117 | 4,257 | 3,863 | 3,633 | 3,482 | 3,374 | 3,293 |
| 10 | 4,965 | 4,103 | 3,708 | 3,478 | 3,326 | 3,217 | 3,136 |
| 11 | 4,844 | 3,982 | 3,587 | 3,357 | 3,204 | 3,095 | 3,012 |
| 12 | 4,747 | 3,885 | 3,490 | 3,259 | 3,106 | 2,996 | 2,913 |
| 13 | 4,667 | 3,806 | 3,411 | 3,179 | 3,025 | 2,915 | 2,832 |
| 14 | 4,600 | 3,739 | 3,344 | 3,112 | 2,958 | 2,848 | 2,764 |
| 15 | 4,543 | 3,682 | 3,287 | 3,056 | 2,901 | 2,791 | 2,707 |
| 16 | 4,494 | 3,634 | 3,239 | 3,007 | 2,852 | 2,741 | 2,657 |
| 17 | 4,451 | 3,592 | 3,197 | 2,965 | 2,810 | 2,699 | 2,614 |
| 18 | 4,414 | 3,555 | 3,160 | 2,928 | 2,773 | 2,661 | 2,577 |
| 19 | 4,381 | 3,522 | 3,127 | 2,895 | 2,740 | 2,628 | 2,544 |
| 20 | 4,351 | 3,493 | 3,098 | 2,866 | 2,711 | 2,599 | 2,514 |
| 21 | 4,325 | 3,467 | 3,073 | 2,840 | 2,685 | 2,573 | 2,488 |
| 22 | 4,301 | 3,443 | 3,049 | 2,817 | 2,661 | 2,549 | 2,464 |
| 23 | 4,279 | 3,422 | 3,028 | 2,796 | 2,640 | 2,528 | 2,442 |
| 24 | 4,260 | 3,403 | 3,009 | 2,776 | 2,621 | 2,508 | 2,423 |
| 25 | 4,242 | 3,385 | 2,991 | 2,759 | 2,603 | 2,490 | 2,405 |
| 26 | 4,225 | 3,369 | 2,975 | 2,743 | 2,587 | 2,474 | 2,388 |
| 27 | 4,210 | 3,354 | 2,960 | 2,728 | 2,572 | 2,459 | 2,373 |
| 28 | 4,196 | 3,340 | 2,947 | 2,714 | 2,558 | 2,445 | 2,359 |
| 29 | 4,183 | 3,328 | 2,934 | 2,701 | 2,545 | 2,432 | 2,346 |
| 30 | 4,171 | 3,316 | 2,922 | 2,690 | 2,534 | 2,421 | 2,334 |
| 40 | 4,085 | 3,232 | 2,839 | 2,606 | 2,450 | 2,336 | 2,249 |
| 60 | 4,001 | 3,150 | 2,758 | 2,525 | 2,368 | 2,254 | 2,167 |
| 120 | 3,920 | 3,072 | 2,680 | 2,447 | 2,290 | 2,175 | 2,087 |
| Inf | 3,842 | 2,996 | 2,605 | 2,372 | 2,214 | 2,099 | 2,010 |

BCU IAŞI "M. EMINESCU" IAŞI

| df2/df1 | 8 | 9 | 10 | 20 | 30 | 120 | INF |
|---|---|---|---|---|---|---|---|
| 1 | 238,883 | 240,543 | 241,882 | 248,013 | 250,095 | 253,253 | 254,314 |
| 2 | 19,371 | 19,385 | 19,396 | 19,446 | 19,462 | 19,487 | 19,496 |
| 3 | 8,845 | 8,812 | 8,786 | 8,660 | 8,617 | 8,549 | 8,526 |
| 4 | 6,041 | 5,999 | 5,964 | 5,803 | 5,746 | 5,658 | 5,628 |
| 5 | 4,818 | 4,773 | 4,735 | 4,558 | 4,496 | 4,399 | 4,365 |
| 6 | 4,147 | 4,099 | 4,060 | 3,874 | 3,808 | 3,705 | 3,669 |
| 7 | 3,726 | 3,677 | 3,637 | 3,445 | 3,376 | 3,267 | 3,230 |
| 8 | 3,438 | 3,388 | 3,347 | 3,150 | 3,079 | 2,967 | 2,928 |
| 9 | 3,230 | 3,179 | 3,137 | 2,937 | 2,864 | 2,748 | 2,707 |
| 10 | 3,072 | 3,020 | 2,978 | 2,774 | 2,700 | 2,580 | 2,538 |
| 11 | 2,948 | 2,896 | 2,854 | 2,646 | 2,571 | 2,448 | 2,405 |
| 12 | 2,849 | 2,796 | 2,753 | 2,544 | 2,466 | 2,341 | 2,296 |
| 13 | 2,767 | 2,714 | 2,671 | 2,459 | 2,380 | 2,252 | 2,206 |
| 14 | 2,699 | 2,646 | 2,602 | 2,388 | 2,308 | 2,178 | 2,131 |
| 15 | 2,641 | 2,588 | 2,544 | 2,328 | 2,247 | 2,114 | 2,066 |
| 16 | 2,591 | 2,538 | 2,494 | 2,276 | 2,194 | 2,059 | 2,010 |
| 17 | 2,548 | 2,494 | 2,450 | 2,230 | 2,148 | 2,011 | 1,960 |
| 18 | 2,510 | 2,456 | 2,412 | 2,191 | 2,107 | 1,968 | 1,917 |
| 19 | 2,477 | 2,423 | 2,378 | 2,156 | 2,071 | 1,930 | 1,878 |
| 20 | 2,447 | 2,393 | 2,348 | 2,124 | 2,039 | 1,896 | 1,843 |
| 21 | 2,421 | 2,366 | 2,321 | 2,096 | 2,010 | 1,866 | 1,812 |
| 22 | 2,397 | 2,342 | 2,297 | 2,071 | 1,984 | 1,838 | 1,783 |
| 23 | 2,375 | 2,320 | 2,275 | 2,048 | 1,961 | 1,813 | 1,757 |
| 24 | 2,355 | 2,300 | 2,255 | 2,027 | 1,939 | 1,790 | 1,733 |
| 25 | 2,337 | 2,282 | 2,237 | 2,008 | 1,919 | 1,768 | 1,711 |
| 26 | 2,321 | 2,266 | 2,220 | 1,990 | 1,901 | 1,749 | 1,691 |
| 27 | 2,305 | 2,250 | 2,204 | 1,974 | 1,884 | 1,731 | 1,672 |
| 28 | 2,291 | 2,236 | 2,190 | 1,959 | 1,869 | 1,714 | 1,654 |
| 29 | 2,278 | 2,223 | 2,177 | 1,945 | 1,854 | 1,698 | 1,638 |
| 30 | 2,266 | 2,211 | 2,165 | 1,932 | 1,841 | 1,684 | 1,622 |
| 40 | 2,180 | 2,124 | 2,077 | 1,839 | 1,744 | 1,577 | 1,509 |
| 60 | 2,097 | 2,040 | 1,993 | 1,748 | 1,649 | 1,467 | 1,389 |
| 120 | 2,016 | 1,959 | 1,911 | 1,659 | 1,554 | 1,352 | 1,254 |
| Inf | 1,938 | 1,880 | 1,831 | 1,571 | 1,459 | 1,221 | 1,000 |

BCU IASI / CENTRAL UNIVERSITY LIBRARY

## Repartiţia Fisher




| Df2/df1 | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|---|
| 1 | 4052,181 | 4999,500 | 5403,352 | 5624,583 | 5763,650 | 5858,986 | 5928,356 |
| 2 | 98,503 | 99,000 | 99,166 | 99,249 | 99,299 | 99,333 | 99,356 |
| 3 | 34,116 | 30,817 | 29,457 | 28,710 | 28,237 | 27,911 | 27,672 |
| 4 | 21,198 | 18,000 | 16,694 | 15,977 | 15,522 | 15,207 | 14,976 |
| 5 | 16,258 | 13,274 | 12,060 | 11,392 | 10,967 | 10,672 | 10,456 |
| 6 | 13,745 | 10,925 | 9,780 | 9,148 | 8,746 | 8,466 | 8,260 |
| 7 | 12,246 | 9,547 | 8,451 | 7,847 | 7,460 | 7,191 | 6,993 |
| 8 | 11,259 | 8,649 | 7,591 | 7,006 | 6,632 | 6,371 | 6,178 |
| 9 | 10,561 | 8,022 | 6,992 | 6,422 | 6,057 | 5,802 | 5,613 |
| 10 | 10,044 | 7,559 | 6,552 | 5,994 | 5,636 | 5,386 | 5,200 |
| 11 | 9,646 | 7,206 | 6,217 | 5,668 | 5,316 | 5,069 | 4,886 |
| 12 | 9,330 | 6,927 | 5,953 | 5,412 | 5,064 | 4,821 | 4,640 |
| 13 | 9,074 | 6,701 | 5,739 | 5,205 | 4,862 | 4,620 | 4,441 |
| 14 | 8,862 | 6,515 | 5,564 | 5,035 | 4,695 | 4,456 | 4,278 |
| 15 | 8,683 | 6,359 | 5,417 | 4,893 | 4,556 | 4,318 | 4,142 |
| 16 | 8,531 | 6,226 | 5,292 | 4,773 | 4,437 | 4,202 | 4,026 |
| 17 | 8,400 | 6,112 | 5,185 | 4,669 | 4,336 | 4,102 | 3,927 |
| 18 | 8,285 | 6,013 | 5,092 | 4,579 | 4,248 | 4,015 | 3,841 |
| 19 | 8,185 | 5,926 | 5,010 | 4,500 | 4,171 | 3,939 | 3,765 |
| 20 | 8,096 | 5,849 | 4,938 | 4,431 | 4,103 | 3,871 | 3,699 |
| 21 | 8,017 | 5,780 | 4,874 | 4,369 | 4,042 | 3,812 | 3,640 |
| 22 | 7,945 | 5,719 | 4,817 | 4,313 | 3,988 | 3,758 | 3,587 |
| 23 | 7,881 | 5,664 | 4,765 | 4,264 | 3,939 | 3,710 | 3,539 |
| 24 | 7,823 | 5,614 | 4,718 | 4,218 | 3,895 | 3,667 | 3,496 |
| 25 | 7,770 | 5,568 | 4,675 | 4,177 | 3,855 | 3,627 | 3,457 |
| 26 | 7,721 | 5,526 | 4,637 | 4,140 | 3,818 | 3,591 | 3,421 |
| 27 | 7,677 | 5,488 | 4,601 | 4,106 | 3,785 | 3,558 | 3,388 |
| 28 | 7,636 | 5,453 | 4,568 | 4,074 | 3,754 | 3,528 | 3,358 |
| 29 | 7,598 | 5,420 | 4,538 | 4,045 | 3,725 | 3,499 | 3,330 |
| 30 | 7,562 | 5,390 | 4,510 | 4,018 | 3,699 | 3,473 | 3,304 |
| 40 | 7,314 | 5,179 | 4,313 | 3,828 | 3,514 | 3,291 | 3,124 |
| 60 | 7,077 | 4,977 | 4,126 | 3,649 | 3,339 | 3,119 | 2,953 |
| 120 | 6,851 | 4,787 | 3,949 | 3,480 | 3,174 | 2,956 | 2,792 |
| Inf | 6,635 | 4,605 | 3,782 | 3,319 | 3,017 | 2,802 | 2,639 |

BCU IAŞI

| Df2/df1 | 8 | 9 | 10 | 20 | 30 | 120 | INF |
|---|---|---|---|---|---|---|---|
| 1 | 5981,070 | 6022,473 | 6055,847 | 6208,730 | 6260,649 | 6339,391 | 6365,864 |
| 2 | 99,374 | 99,388 | 99,399 | 99,449 | 99,466 | 99,491 | 99,499 |
| 3 | 27,489 | 27,345 | 27,229 | 26,690 | 26,505 | 26,221 | 26,125 |
| 4 | 14,799 | 14,659 | 14,546 | 14,020 | 13,838 | 13,558 | 13,463 |
| 5 | 10,289 | 10,158 | 10,051 | 9,553 | 9,379 | 9,112 | 9,020 |
| 6 | 8,102 | 7,976 | 7,874 | 7,396 | 7,229 | 6,969 | 6,880 |
| 7 | 6,840 | 6,719 | 6,620 | 6,155 | 5,992 | 5,737 | 5,650 |
| 8 | 6,029 | 5,911 | 5,814 | 5,359 | 5,198 | 4,946 | 4,859 |
| 9 | 5,467 | 5,351 | 5,257 | 4,808 | 4,649 | 4,398 | 4,311 |
| 10 | 5,057 | 4,942 | 4,849 | 4,405 | 4,247 | 3,996 | 3,909 |
| 11 | 4,744 | 4,632 | 4,539 | 4,099 | 3,941 | 3,690 | 3,602 |
| 12 | 4,499 | 4,388 | 4,296 | 3,858 | 3,701 | 3,449 | 3,361 |
| 13 | 4,302 | 4,191 | 4,100 | 3,665 | 3,507 | 3,255 | 3,165 |
| 14 | 4,140 | 4,030 | 3,939 | 3,505 | 3,348 | 3,094 | 3,004 |
| 15 | 4,004 | 3,895 | 3,805 | 3,372 | 3,214 | 2,959 | 2,868 |
| 16 | 3,890 | 3,780 | 3,691 | 3,259 | 3,101 | 2,845 | 2,753 |
| 17 | 3,791 | 3,682 | 3,593 | 3,162 | 3,003 | 2,746 | 2,653 |
| 18 | 3,705 | 3,597 | 3,508 | 3,077 | 2,919 | 2,660 | 2,566 |
| 19 | 3,631 | 3,523 | 3,434 | 3,003 | 2,844 | 2,584 | 2,489 |
| 20 | 3,564 | 3,457 | 3,368 | 2,938 | 2,778 | 2,517 | 2,421 |
| 21 | 3,506 | 3,398 | 3,310 | 2,880 | 2,720 | 2,457 | 2,360 |
| 22 | 3,453 | 3,346 | 3,258 | 2,827 | 2,667 | 2,403 | 2,305 |
| 23 | 3,406 | 3,299 | 3,211 | 2,781 | 2,620 | 2,354 | 2,256 |
| 24 | 3,363 | 3,256 | 3,168 | 2,738 | 2,577 | 2,310 | 2,211 |
| 25 | 3,324 | 3,217 | 3,129 | 2,699 | 2,538 | 2,270 | 2,169 |
| 26 | 3,288 | 3,182 | 3,094 | 2,664 | 2,503 | 2,233 | 2,131 |
| 27 | 3,256 | 3,149 | 3,062 | 2,632 | 2,470 | 2,198 | 2,097 |
| 28 | 3,226 | 3,120 | 3,032 | 2,602 | 2,440 | 2,167 | 2,064 |
| 29 | 3,198 | 3,092 | 3,005 | 2,574 | 2,412 | 2,138 | 2,034 |
| 30 | 3,173 | 3,067 | 2,979 | 2,549 | 2,386 | 2,111 | 2,006 |
| 40 | 2,993 | 2,888 | 2,801 | 2,369 | 2,203 | 1,917 | 1,805 |
| 60 | 2,823 | 2,718 | 2,632 | 2,198 | 2,028 | 1,726 | 1,601 |
| 120 | 2,663 | 2,559 | 2,472 | 2,035 | 1,860 | 1,533 | 1,381 |
| inf | 2,511 | 2,407 | 2,321 | 1,878 | 1,696 | 1,325 | 1,000 |

BCU IASI/CENTRAL UNIVERSITY LIBRARY

Culoarea este o pasiune, iar pasiunea se împarte...

tipografia

sedcom libris

TIPOGRAFIA SEDCOM LIBRIS
IAȘI, ROMÂNIA
ISO 9001:2000
Șos. Moara de Foc nr. 4,
cod 700572, Iași, România
Tel.: +40.232.234.582, 270.479, 242.877;
Fax.: +40.232.233.080
E-mail: tipografia@sedcom.ro
http:// www.sedcom.ro

Pentru că ne pasă cine ești, ce faci, cum faci și unde vrei să ajungi...

BCU IASI/CENTRAL UNIVERSITY LIBRARY

BCU IASI/CENTRAL UNIVERSITY LIBRARY

I.S.B.N.: (10) 973-670-208-1
I.S.B.N.: (13) 978-973-670-208-2